35

158
12.8.70

1382
112

Psychological Monographs:

General and Applied

Combining the *Applied Psychology Monographs* and the *Archives of Psychology*
with the *Psychological Monographs*

VOLUME 67

1953

Editor

HERBERT S. CONRAD
Department of Health, Education, and Welfare
Office of Education
Washington 25, D.C.

Consulting Editors

DONALD E. BAIER
FRANK A. BEACH
ROBERT G. BERNREUTER
WILLIAM A. BROWNELL
HAROLD E. BURTT
JERRY W. CARTER, JR.
CLYDE H. COOMBS
JOHN G. DARLEY
JOHN F. DASHIELL
EUGENIA HANFMANN
EDNA HEIDBREDER
HAROLD E. JONES
DONALD W. MACKINNON
LORRIN A. RIGGS
CARL R. ROGERS
SAUL ROSENZWEIG
ROSS STAGNER
PERCIVAL M. SYMONDS
JOSEPH TIFFIN
LEDYARD R TUCKER
JOSEPH ZUBIN

Published by
THE AMERICAN PSYCHOLOGICAL ASSOCIATION
1333 SIXTEENTH STREET N.W., WASHINGTON 6, D.C.

Bureau Ednl. Psy. Research
DAVID HARE TRAINING COLLEGE
Dated ……………………
Accs No ……………………

[illegible] Research
12.8.70
J 158

CONTENTS OF VOLUME 67

Whole No.

351 Personality Determinants of Vocational Choice. Leonard Small. Pp. 21.

352 A Psychological Study of Eminent Psychologists and Anthropologists, and a Comparison with Biological and Physical Scientists. Anne Roe. Pp. 55.

353 The Development of a Test for the Measurement of Anxiety: a Study of Its Reliability and Validity. M. J. Freeman. Pp. 18.

354 Traumatic Avoidance Learning: Acquisition in Normal Dogs. Richard L. Solomon and Lyman C. Wynne. Pp. 19.

355 Errors in Time-Study Judgments of Industrial Work Pace. Kalman A. Lifson. Pp. 14.

356 A Quantification, Standardization, and Validation of the Bender-Gestalt Test on Normal and Neurotic Adults. Wallace Gobetz. Pp. 28.

357 Psychological Changes during the First Year following Prefrontal Lobotomy. I. W. Scherer, J. F. Winne, D. D. Clancy, and R. W. Baker. Pp. 24.

358 Some Physiological Concomitants of Mental Work. Edward W. Geldreich. Pp. 28.

359 A Model of the Auditory Threshold and Its Application to the Problem of the Multiple Observer. Moncrieff Smith and Edna A. Wilson. Pp. 35.

360 Passivity in a Class of Peptic Ulcer Patients. Alvin Scodel. Pp. 15.

361 Democratic Leadership in the College Classroom. Donald M. Johnson and Henry Clay Smith. Pp. 20.

362 The Prediction of Effectiveness as a Factory Foreman. Martin M. Bruce. Pp. 17.

363 Responses to Auditory Stimuli at the Cochlea and at the Auditory Cortex. Mark R. Rosenzweig and Walter A. Rosenblith. Pp. 26.

364 Job Satisfaction as Related to Need Satisfaction in Work. Robert H. Schaffer. Pp. 29.

365 The Effects of Analeptic Drugs in Relieving Fatigue. Robert H. Seashore and A. C. Ivy. Pp. 16.

366 Situational and Personality Factors in Leadership among Sorority Women. Bernard M. Bass, Cecil R. Wurster, Paddy Ann Doll, and Dean J. Clair. Pp. 23.

367 The Development of Phobias in Married Women. A. Stanley Webster. Pp. 18.

368 Assessing Social-emotional Climate in the Classroom by Withall's Technique. Harold E. Mitzel and William Rabinowitz. Pp. 19.

369 An Experimental Study of Extinction. Marshall B. Jones. Pp. 17.

370. Vocabulary Knowledge and Usage among Normal and Schizophrenic Subjects. Louis J. Moran. Pp. 19.

371. The Use of Modified Block Designs in the Evaluation and Training of the Brain-injured. Marion Reissenweber. Pp. 28.

738L1

Vol. 67, No. 1 Whole No. 351, 1953

Psychological Monographs: General and Applied

Personality Determinants of Vocational Choice[1]

Leonard Small
Vocational Advisory Service

I. THE PROBLEM AND THE POPULATION STUDIED

What personality factors determine the choice of a vocation? The investigation reported here has sought some of the answers to this question.

Problem. We decided to compare the vocational choices made by adjusted adolescent boys with those of disturbed adolescent boys and to study the reality- and fantasy-contents of their choices.

Assumptions. Our major assumption was that people seek satisfaction of their basic needs in every major aspect of their lives—including the vocational. When a need is consciously felt, it is subject to control and modification by realistic factors. When it is unconscious, our control is less direct and often tenuous. The person with a healthy ego is in strong contact with reality. A healthy ego permits the individual to check his drive for need-satisfaction against the facts provided by the environment and his own faculties. A healthy ego is also willing to postpone satisfactions for a reasonable purpose, and is unlikely to seek a satisfaction in a situation which holds no promise of yielding it. The healthy ego is realistic.

This does not mean that the healthy ego does not experience fantasy—the wishful satisfaction of needs. Fantasy is common property, available to the well and ill alike. The difference rests in the nature of the individual's action: is his action dominated by fantasy or is his fantasy tempered by reality? From these thoughts our assumptions were formulated: (*a*) a difference between "disturbed" and "better-adjusted" individuals is to be found in the strength of their respective egos; (*b*) vocational choice is in good part a function of the ego; (*c*) people seek satisfaction of their needs in their vocations; (*d*) vocational choices are also based upon reality considerations; (*e*) vocational choices require a compromise between fantasy and reality.

Hypothesis. Does it therefore follow that adolescents with emotional disturbances are less realistic in their vocational choices than are better-adjusted adolescents? We arrived at the following hypothesis: Individuals with different ego strengths will show differences in the use they make of reality and wishful fantasy in making their vocational choices.

---

[1] This investigation was supported by a research grant from the National Institute of Mental Health, of the National Institutes of Health, Public Health Service.

The research and this publication were accomplished through the cooperative efforts of Jane F. Culbert, Peter H. Stevens, Dr. John J. Vetter, Anne W. Lloyd, Dr. Jack T. Huber, William F. Marvin, John T. Roberts, Boris von Arnold, Nancy D. Larson, Grace D. Treat, Emma Seipp, Carol A. Rivlin, Betty L. Lewis, and Verna Small.

Thanks are owed to Dr. Helen R. Smith and Mrs. Lois Ruml for their interest and sponsorship.

Specific problems. The formulation of subproblems emerged directly from our concern with this interaction of reality and fantasy. We directed our attention separately to reality and fantasy:

1. Can the reality content of a vocational choice be identified and evaluated?
2. Can the fantasy content of a vocational choice be identified and evaluated?

We then evolved methods for seeking answers to these questions.

The study population. To test the hypothesis we observed adolescent boys. Setting the age limits from 15 to 19 assured us adolescents in a variety of school and work situations.

Having postulated vocational choice to be an ego function, we selected boys who differed in ego strength and set up two groups: (*a*) a definitely maladjusted group who came to the Vocational Advisory Service from psychiatric hospitals and clinics and whom we called the Disturbed Group; (*b*) a better-adjusted group who came to us from schools and similar sources and whom we called the Comparison Group. In each group we selected ten boys at each age level: 15, 16, 17, 18, and 19. The ten boys in each age group consisted of eight white and two Negro boys. The religious distribution at each age level was five Catholic, three Protestant, and two Jewish boys. All the boys were native-born Americans of families in the lower-middle-class economic group. The racial and religious distributions were based upon our knowledge of the distribution of these factors among the disturbed boys referred to our agency in the past.

The matching of subjects in the Disturbed and Comparison Groups was based on age, religion, race, and economic status of the family. We did not expect that educational or intelligence level could be used as equating factors since these two measures are depressed functionally by emotional disturbance (11). Thus we found the mean IQ of the Comparison Group to be 112.1, with an *SD* of 13.6, and that of the Disturbed Group to be 102.2, with an *SD* of 14.1. Intelligence was measured by the Wechsler-Bellevue.

The histories of the disturbed subjects assured their belonging in the Disturbed Group—histories of psychiatric hospitalizations or treatment, of witnessed behavior clearly indicating emotional disturbance. No effort was made to limit these subjects diagnostically. Accordingly we have included neurotics, psychotics, behavior disorders, and delinquents—unquestionably a disturbed group. Selecting the subjects for the Comparison Group was a more difficult procedure. The data accompanying registration for counseling was meager for this purpose; all too often a boy would emerge as a disturbed personality only after counseling was well under way and research data had been obtained from him. We could not ask these clients coming for vocational counseling to submit to a psychiatric interview, first, because that would alter the nature of our relationship with them and, second, because we were afraid we would lose many if we ventured uninvited into this area. We settled for an examination of case histories and counseling record by a psychiatrist.

II. THE REALITY CONTENT OF VOCATIONAL CHOICE

Each subject was asked to name five vocational choices in order of preference. We later limited ourselves to the first two choices because many boys would make only two choices even when pressed for more; the third, fourth, and fifth choices were usually minor variations of the first two choices.

A. Ratings of Realism

Each choice was coded according to the job classifications in Part IV of the *Dictionary of Occupational Titles* of the U. S. Employment Service. These "Entry Classifications" offer a method for relating the capacities of young beginning workers to general fields of work—those in which they subsequently may be expected to progress to the full skills required by specific jobs. This volume lists the personal characteristics found important for entry into each general occupational field. We adopted these characteristics with some very minor modifications, called them Job Requirements, and applied them to each job chosen by a boy. Each listed characteristic was assigned a value of 3 on a four-point (0-3) rating scale. Using the same characteristics and scale, we then rated each subject and called these ratings Personal Characteristics. These ratings were done by the counselor with whom the boy was working. The counselor was in the best position to make Personal Characteristic ratings: he was an objectively minded person with close acquaintance with the boy, and had for some time been considering the boy in these very terms—his suitability for various general fields of work.

The two ratings—Job Requirements and Personal Characteristics—were then compared and the discrepancy between them totaled for each boy. The sum of the discrepancy (hereafter called the Reality-Deviation Score) constitutes a numerical evaluation of the reality of the boy's choice: the smaller the discrepancy value the closer to reality is the choice. Reality-Deviations were scored separately for the first choice (RD_1) and the second choice (RD_2). The second vocational choice of some subjects fell into the same entry classification as their first choice. Since these received the same Reality-Deviation Scores for both choices, they are included only in the RD_1 findings and not in the RD_2 findings.

Table 1 shows a significant difference in the reality contents of the first vocational choice of the Comparison and Disturbed Groups. The better-adjusted boys composing the Comparison Group deviate less from the requirements of their chosen occupations than do the disturbed boys. The first choices of the members of the Comparison Group are more realistic than are those of the disturbed boys.

The difference in RD_2 follows the same trend. Although the Comparison Group is again more realistic than the Disturbed Group in the second choice, the statistical significance of the difference is not as great as it is in the first choice. Comparison of the mean RD_1 and RD_2 values for

TABLE 1

Mean Reality-Deviation Scores for the First and Second Vocational Choices of the Comparison and Disturbed Groups

Group	Datum	RD_1	RD_2
Comparison	Mean	15.3	21.8
	SD	10.0	10.9
	N	50	30
Disturbed	Mean	29.4	26.2
	SD	13.4	12.2
	N	50	38
Critical Ratio		8.8	1.6

the two groups separately shows that the better-adjusted subjects' second choice is less realistic than their first choice, while the second choice of the disturbed subjects is more realistic than their first choice.

We had observed this tendency as we interviewed and tested the boys. We believe that these differences emphasize the dynamics of ego balance. Contact with reality is an active relationship in which reality perceptions are constantly being integrated with fantasy or needs. Compensatory balances are maintained. The better-adjusted boys throw the full weight of their reality perceptions into their first choice—the occupational selection that means most to them. This done, they permit the fantasy components a greater role in making the second choice. The disturbed boys also show the shift of operating forces but with a different emphasis. By inference the greater the departure from reality, the stronger is the operation of fantasy. Fantasy predominates the first choices in the disturbed, and yields somewhat to reality in the second choices.

B. The Developmental Theory of Vocational Choice

After this research began, Ginzberg and his associates (5) published a contribution to a theory of occupational choice based on 64 subjects. They hold that the determinant of choice is predominantly fantasy before the age of 11 and progressively swings toward realistic factors as the child grows older.

We were able to test the application of this theory to our population because we had 100 subjects, including equal numbers of boys at the ages of 15-19.

No evidence was found supporting the developmental theory. Instead of a linear relationship between age and reality, something different appears to take place in both groups at each age. The change is sometimes toward greater realism and at other times to less realism. The absence of a definitive trend supports our belief that the maintenance of reality perception is an active one involving compensatory balances.

C. Other Differences Between the Groups

As noted earlier, the first and second choices of some subjects fell into the same entry classification. Almost twice as many Comparison Group subjects ($N = 20$) made such similar choices as did Disturbed Group subjects ($N = 12$). We paid little attention to this fact until we noted that the IQ of the subjects making similar choices was higher than those whose first and second choices fell into different entry classifications.

A first thought is that this difference is purely a function of intelligence, but we are dealing here with differences in intelligence as measured by tests. That is, the lower IQ of the Disturbed Group is not necessarily the result of a difference in basic native intelligence but could reflect impairment in the use of their endowment (i.e., their relative ego weakness).

We turned our study in this direction. What IQ differences are there between those subjects whose choices are similar and those whose choices are different? We studied as well the differences in the mean subtest scatter on the Wechsler-Bellevue. As third and fourth factors we examined the mean scores of reality deviation for the first and second choices. The results of this study are tabulated in Tables 2 and 3.

The differences found are not signifi-

TABLE 2

DIFFERENCES IN THE COMPARISON GROUP IN ENTRY CLASSIFICATION, IQ, WECHSLER-BELLEVUE SCATTER, AND REALITY-DEVIATION SCORES

Entry Classification of 1st and 2nd Choices	N	Mean Age	Datum	IQ	Mean W-B Scatter	RD_1	RD_2
Same	20	17	Mean	121.3	1.79	12.0	
			SD	9.3	.50	9.1	
Different	30	17	Mean	107.5	1.99	17.5	21.8
			SD	13.1	.51	14.8	10.9
			Critical Ratio	4.4	1.4	1.5	

cant by statistical criteria. The data, however, are provocative. The trend suggests a positive relationship between similar vocational choices on the one hand, with higher IQ, less functional IQ scatter, and greater realism, on the other. A warranted assumption from clinical evidence is that acceptance of reality and greater functional intelligence are ego traits. The assumption would be extended to include the making of similar vocational choices as an ego function. Nearly twice as many of the better-adjusted subjects make similar choices than do disturbed subjects; their IQ is higher; they perform more homogenously on the Wechsler-Bellevue; and their choices are more realistic.

The making of similar vocational choices suggests an economy in self-perception: The stronger ego sees more clearly the limits to what the individual can do successfully. We shall return to this topic in our discussion.

TABLE 3

DIFFERENCES IN THE DISTURBED GROUP IN ENTRY CLASSIFICATION, IQ, WECHSLER-BELLEVUE SCATTER, AND REALITY-DEVIATION SCORES

Entry Classification of 1st and 2nd Choices	N	Mean Age	Datum	IQ	Mean W-B Scatter	RD_1	RD_2
Same	12	17	Mean	103.9	1.86	30.8	
			SD	14.1	.44	8.7	
Different	38	17	Mean	101.9	2.13	29.0	26.2
			SD	15.2	.75	14.6	12.2
			Critical Ratio	.4	1.6	.5	

III. THE FANTASY CONTENT OF VOCATIONAL CHOICE

It will be recalled that the second choice of the disturbed boy tends to be, on the average, more realistic than his first choice, while the second choice of the better-adjusted boy is, on the average, less realistic than his first choice. Apparently something operates in conjunction with or in opposition to realism in the making of choices. We postulated that this other factor is the emotional needs or fantasies of the individual.

To evaluate the fantasy content of the choice we devised a special interview—the Job-Concept interview[2]—to elicit fantasy about jobs. The Job-Concept interview is divided into the fantasy involved in the first choice and that involved in the second choice. These parts were rated separately on a three-point scale for each of 27 needs derived from the work of Murray (9).[3]

To check the Job-Concept findings we administered 15 cards of the Thematic Apperception Test and rated the responses on the same scale. In a similar effort we rated the Case History of each individual on the same scale. These additional approaches, however, were not very efficient. We found that the Job-Concept interview method was more sensitive to specific needs than either the Thematic Apperception Test or Case History methods. Only the data from the Job-Concept interview method will be reported here in detail. We refer to the ratings respectively as:

1. First Choice Job-Concept Needs—JC_1.
2. Second Choice Job-Concept Needs—JC_2.

A. THE RELIABILITY OF THE RATINGS

All data were rated by three independent judges: a counselor, the research psychologist, and the principal investigator. The counselors rated only the data of the boys they counseled, while the research psychologist and the principal investigator rated all of the boys.

The chi-square test was applied to measure the extent of agreement between judges. Ratings within .5 of each other were said to agree. Yates's correction was used in computing all chi-square values because it is slightly more conservative than the direct method. We found the critical point to be 22 agreements, which produced a chi-square value of 4.66 and a p value falling between .05 and .01.

Agreement was measured separately for the ratings of JC_1 and JC_2. The research psychologist and the principal investigator agreed 200 times in 200. The other judges (counselors) seldom agreed with the research-staff judges *or* with each other.

The reliability of the method is based upon the agreement (200 times in 200) of the research-staff judges who lived more closely with the concepts involved in the research than the counselors and had more experience with projective techniques and with the concepts of fantasy productions.

B. THE VALIDITY OF THE METHOD

The rating scores of all subjects for each need (to account for both frequency and intensity) were totalled separately for each of the three methods. The needs obtained by each of the methods were then ranked, permitting rank-order correlations between the three methods. This procedure produced a coefficient of +.44 between the Job-Concept interview and the Thematic Apperception Test methods, and +.52 between the Job-Concept interview and Case History methods.

We conclude that the Job-Concept interview method has a capacity to assay the relative frequency and intensity of needs that is positively related to the capacity of both the Thematic Apperception Test and Case History methods to make similar assessments.

---

[2] See IV, A for description and illustration.

[3] See Appendix A for the list of needs and their definitions.

C. The Fantasy Content of the Choices

To simplify the handling of the data we used the Job-Concept ratings of one of the two judges most often in agreement. We assigned a value of 0 to a rating of "None" or "Little," 1 to "Some," and 2 to "Great." Group totals were calculated for the Comparison and Disturbed Groups separately for each need. The mean of the differences between the two groups on *all needs* was used as a guide to the significance of differences between specific needs. The results of this study are found in Table 4.

Comparison of First-Choice Job-Concept Needs (JC_1) and Second-Choice Job-Concept Needs (JC_2) shows differences between the two groups of subjects which vary in degree of significance. The Comparison Group exceeds the Disturbed Group significantly (the Mean Difference or more in both JC's) in Order, Achievement, Recognition, and Affiliation. Less significantly (the Mean Difference or more in one JC and less than one Mean Difference in the other JC) the Comparison Group exceeds the Disturbed Group in Dominance, Deference, and Sympathy.

The Disturbed Group exceeds the Comparison Group significantly in Inward Pain and Tension Discharge, and less significantly in Avoidance.

In other needs one group consistently exceeds the other, but without any degree of significance (the differences in both JC's being less than the Mean Difference). In this manner the Comparison Group shows more Conservance, Retention, Attack, Autonomy, Uniqueness, and Tension Accumulation, and the Disturbed Group shows more Removal.

It is striking that both Deference and Dominance characterize the Comparison Group. This seeming contradiction is not inexplicable. It reflects both the desire for leadership and the awareness by the subject of his inexperience and need to learn—evidence of a healthy ego, functioning to integrate wish and reality.

In handling Humiliation the Comparison Group more frequently elects to attack the source of humiliation by striving and by solution of the more difficult task. The Disturbed Group avoids activity that they suspect may be beyond their power. This is corroborated in the Affection group of needs. The Disturbed Group is characterized by the Removal need while more members of the Comparison Group seek the opportunity for Affiliation and to exercise Sympathy.

In the handling of Tension there is repeated the difference in relationship to the environment. The Disturbed Group shows more Tension Discharge (to act out their tension) while the Comparison Group, seemingly more concerned with maintaining a rapprochement with the environment, restrains this acting out and shows more Tension Accumulation. Finally, the Disturbed Group shows more of a tendency toward masochism in the Inward Pain need.

These differences reflect a difference in ego function. The members of the Comparison Group exercise their skills, talents, and intelligence in seeking an involvement with the environment. The members of the Disturbed Group turn away from the environment and in so doing restrict their egos.

Thus far we have accounted for only 18 of the 27 needs; in the remaining 9 there is observed a phenomenon we have named "reversal." An example of reversal is noted in Emulation. The Disturbed Group shows more of this need in JC_1, while in JC_2 it is the Comparison Group that shows more of the need.

In the First-Choice Job Concept the Comparison Group shows more Acquisition, Impulse Restraint, and Asking, but loses this leadership in the Second-Choice Job Concept to the Disturbed Group. Conversely, in the first JC the Disturbed Group shows more Emulation, Impulse Release, Exhibition, Justification, Outward Pain, and Dependence, but yields

TABLE 4

DIFFERENCES IN TOTAL RATING SCORES BETWEEN THE COMPARISON GROUP (C) AND THE DISTURBED GROUP (D) ON EACH NEED AS OBTAINED FROM THE JOB-CONCEPT INTERVIEW

		Total Scores JC_1		Difference	Total Scores JC_2		Difference
		C	D		C	D	
OBJECTS—	ACQUISITION	31	22	+ 9	19	22	3
	CONSERVANCE	12	9	3	15	12	3
	ORDER	27	21	+ 6	33	24	+ 9
	RETENTION	11	7	4	5	5	0
AMBITION—	ACHIEVEMENT	29	20	+ 9	26	12	+14
	RECOGNITION	47	21	++26	45	27	++18
	EXHIBITION	24	25	1	16	9	7
HUMILIATION—	AVOIDANCE	32	33	1	27	37	−10
	JUSTIFICATION	2	3	1	3	1	2
	ATTACK	9	4	5	5	3	2
POWER—	DOMINANCE	41	35	+ 6	40	35	5
	DEFERENCE	24	18	+ 6	19	12	7
	EMULATION	26	33	− 7	29	21	+ 8
	AUTONOMY	30	26	4	27	22	5
	UNIQUENESS	7	6	1	10	6	4
PAIN—	OUTWARD	36	40	4	34	31	3
	INWARD	51	65	−−14	48	59	−11
IMPULSES—	RELEASE	11	16	5	19	8	+11
	RESTRAINT	34	30	4	23	37	−14
AFFECTION—	AFFILIATION	26	16	+10	26	13	+13
	REMOVAL	13	17	4	13	17	4
	SYMPATHY	37	28	+ 9	40	34	6
	DEPENDENCE	29	33	4	33	27	6
TENSION—	DISCHARGE	52	66	−−14	44	58	−14
	ACCUMULATE	6	3	3	3	3	0
INFORMATION—	ASKING	25	12	++13	22	26	4
	TELLING	6	6	0	10	2	+ 8
MEAN DIFFERENCE				6.4			8.4

+ = C > D by once to twice the Mean Difference.
− = C < D by once to twice the Mean Difference.
++ = C > D by twice or more the Mean Difference.
−− = C < D by twice or more the Mean Difference.

this primacy in the second JC to the Comparison Group.

The reversals throw some light upon the phenomenon noted in the Reality-Deviation Scores. It will be recalled that the Comparison Group tended to become less realistic in their second choice, while the Disturbed Group became more realistic in their second choice. The shift to less reality in the Comparison Group is accompanied by an increase in acting out and environment-avoidance needs, while the shift to greater reality in the Disturbed Group is accompanied by an increase of environment-involvement needs.

Greater realism in vocational choices, therefore, is found in the individual whose ego successfully mediates against the emergence of egocentric and asocial impulses, and instead directs the personality towards relationships with objects and people, status, independence, and the exercise of talents and skills. Less realism in vocational choice is found in the individual whose ego is restricted and less successful in mediating asocial, withdrawal, and aggressive needs.

1. *Agreement between Needs of First- and Second-Choice Job Concepts*

We applied the chi-square test to ratings of needs revealed in the Job Concepts of the first and second choices of each subject to obtain a measure of the similarity or diversity that might exist in the fantasy content of the two choices.

Reference to Table 5 shows that more of the Disturbed Group show a significant agreement between the needs expressed in the Job Concepts of their two choices. Such agreement would be expected where the choices are similar. But the Disturbed Group made fewer similar choices than did the Comparison Group. The Disturbed Group more often seeks satisfaction of similar needs in dissimilar situations. The investigation of this point was extended in two ways. We first calculated the percentage of subjects in each group showing significant agreement between the needs expressed in the Job Concepts of the two choices according to whether their choices were in the same or different entry-classification areas. Secondly, we calculated the mean number of differences between the needs expressed in the Job Concepts of the two choices according to the same grouping.

TABLE 5

FREQUENCY OF SIGNIFICANT AGREEMENT IN NEEDS BETWEEN FIRST-CHOICE AND SECOND-CHOICE JOB-CONCEPTS OF THE TWO GROUPS

Group	*p* Value		Totals
	Below .05	Above .05	
Comparison	14	36	50
Disturbed	22	28	50

Table 6 presents the results of these calculations. The percentage column shows that of those whose choices are in the same entry classification, 50 per cent of the adjusted boys show significant agreement between the fantasy content of their two choices. Only 33 per cent of the disturbed boys show such agreement. Among those subjects whose choices fall in different classifications, 13 per cent of the Comparison Group and 45 per cent of the disturbed boys show significant agreement.

When members of the Comparison Group made similar vocational choices, more of them sought the satisfaction of similar needs in both vocations. Where members of the Comparison Group made unlike vocational choices, fewer of them

TABLE 6

PERCENTAGE OF SUBJECTS IN SAME AND DIFFERENT ENTRY CLASSIFICATIONS SHOWING SIGNIFICANT AGREEMENT IN NEEDS EXPRESSED IN JC_1 AND JC_2, AND THE MEAN NUMBER OF DIFFERENCES IN NEEDS FOR THE SAME GROUPS

Subjects	Entry Classification	Percentage Showing Significant Agreement between JC_1 and JC_2	Mean Number of Differences between JC_1 and JC_2
Comparison	Same	50	6.5
	Different	13	7.9
Disturbed	Same	33	6.3
	Different	45	6.1

sought the satisfaction of similar needs in both vocations. Fewer of the Disturbed Group sought satisfaction of similar needs in like vocations and more of them sought satisfaction of similar needs in unlike vocations.

This interpretation is reinforced by the column in Table 6 which shows the mean number of differences between the fantasy content of the two choices. The Disturbed Group members whose choices are unlike show the *least* number of differences while the largest number of differences is shown by the Comparison Group members whose choices are unlike. This supports the conclusion that the Comparison Group members more often seek satisfaction of similar needs when the realistic situation is likely to provide them. The Disturbed Group shows a pronounced tendency to seek satisfaction of the same needs even when the realistic situation is less likely to provide them. Again we find a difference between the two groups in reality perception and in ego mediation of needs.

IV. EXAMPLES OF JOB-CONCEPT FANTASIES

As the ratings of the Job-Concept interviews were made, illustrations of specific needs emerging in job fantasy were noted, as were evidences of patterning. Examples of these needs are presented here, along with a general description of the Job-Concept interview and the record of an interview around the first vocational choice of a better-adjusted boy.

A. THE JOB-CONCEPT INTERVIEW

The Job-Concept interview was designed to elicit the fantasy content of the individual's job thinking. His vocational aversions and reasons for them are evoked. Identifications are explored. The individual's comprehension of the realities involved in his vocational choice is studied—his concept of what is done on the job and how. He is then led into an expression of deeper fantasy, by being asked to describe a future typical day of employment in his chosen work and to tell what he believes to be the attitudes of family, friends, society, and self to those in the occupation. The "typical day" and "attitudes" become projections of himself. Finally, to reach a still deeper level, we ask for the subject's actual or made-up dreams about his vocational choices.

The following is the record of the Job-Concept interview with a better-adjusted boy about his first choice:

1. *Aversions*

a. Armed forces. Too regular and strict. Don't allow for human feelings. I had a cousin in the Army and a few of my mother's friends were. The work isn't unpleasant but it's mostly the regulation.

b. Policeman. They have to direct traffic. Their general policy to most people is that they are crooks or pests. Directing traffic is a mean job, standing around.

c. Elevator operator. I'd be wasting time on that. There are better opportunities than doing that.

d. Clerk. Same as elevator operator.

e. Painter. I don't really hate this but it is like a lot of jobs—no advancement and don't acquire any skills that can be beneficial to you.

2. *First choice—Forester*

a. Reasons for selecting. I like the country so much better. It's healthier. I like plants and animals, say nature in general. I was very interested in science. There are lots of fields in forestry that use science.

b. Identifications. I don't know anyone directly. A schoolmate of my mother's has three sons and the oldest is in forestry. I haven't met him but his mother has written a few letters about him. He's married now but besides that, I figure he's interested in agriculture and he's lived in an agricultural place. He's young yet and likes to kid around. The things I like about his work are that I heard about his being in college—since then I've read quite a few books and pamphlets about forestry. I decided on forestry about a year and a half ago. I've known about him for a long time. He went through the course fast.

Nobody has encouraged me but my mother says if that's what I like that I should read about it.

c. Job concept. (What is done?) He's a combination—has to have the same training as a ranger that includes animal care, plant care, fire fighting and prevention. The forester has to be a director of other men and tell them how to put out a fire and figure how many more men he'll need or to plan for next season. They already have the policy but have to figure how to handle vandalism and prevent fires.

(How is it done?) He works from experience, as in the fire fighting. He sees what kind of fire it is. Figures scientific points, the way the wind is blowing, dryness of area, how many bodies of water there are and what type are in that area. Then he figures how many foresters he has in that area. Depending on these he figures how many more men and how much machinery he'll need for fire fighting.

If they have to figure animal conservation, he locates hunting areas in his sector and posts warnings for hunters and orders scarce animals brought in so he can breed them, and orders food for those that can't find their own in the winter.

It is his general duty to gather daily reports of the rangers in his area—checking on their supplies and equipment. And he has to make reports to the government.

The part that would appeal most to me is animal and plant care.

(Describe typical day) Have breakfast, check reports of all the other rangers. I'd read over reports of how many animals there were and say (smiling), "Section 14 did good." I'd need about 15 bales of hay for the next two weeks. Then I'd ride down to the south to the deer pens and ask the ranger how much food he'd need for the deer and when he could use some new fawns. After that I'd go to the intake station—this is a national park—and I'd find the intake of tourists. When the road patrol came in I'd get the names or the license numbers of any of the tourists who were troublemakers and have them posted at all outlet stations. Then I'd have to go to Section 7 and take a sample of the fungus that was getting the beech trees in that area. On the way back to my laboratory I'd pick up a broken pump that had to be fixed and take it to the shop and drop my fungus specimen off at the lab and go home. After dinner I'd take in evening calls from the rangers and make out my daily reports. I've never been to a national park but I've read a lot so I have a good basis for it.

d. Attitudes of others. (What do people think?) They have a poor conception of it. They see men in uniform riding a horse giving them regulations. Some think it's romantical and others think it's a simple job. They're respected by the average citizen but mostly they are thought of as country cops.

(What do family think?) His wife is probably just as concerned in it as he is. Not only for his pay check but because fire and plagues may come into their lives.

Since his parents won't see much of him they wouldn't belittle it but they wouldn't think too much of it. His children might be interested in learning plant care and animal breeding because they like pets. I have about seven cats at home.

(What do his friends think?) The same as he does, for most of his friends would be foresters.

(What would a forester think?) He's proud because it is important and because he is interested in it.

e. Make up dream. When you dream, you always dream of what you'd like to happen. So I'll say a fire because that's the most important thing. I find there's a fire in the north section and know the wood is dry because the rainy season is just about to come. I get the wind speed and chances for precipitation and judge how many men I have to get up there and how fast. Then I get two airplanes. My men are all jumpers. While I'm going to the fire, I order by short wave two bulldozers and 100 axes sent to the scene. The men jump and I do too. After we cut trees to make a fire bridge, we set a backfire and we have to get out fast. There is one tall pine tree that may cross the guys so I order a bulldozer to push it down. We have to rush around to the side because of a shift in the wind but one of my men had foreseen that possibility and took care of it. The next day we have to estimate how much damage was done and how soon trees can be planted there.

The needs expressed in the Aversions relate to Automy and Achievement. The boy shows a remote Emulation, considerable Information Asking, some Telling, Dominance, Order, and some Inward Pain. A striking need is to show Sympathy in the care of animals and plants. He shows a concern with aggression expressed in his thoughts about vandalism and fire fighting. He anticipates fire and plague (a masochistic fantasy) but also wishes to fight them (an aggressive fantasy). Following the destruction he replants and nurtures (Sympathy).

B. EXAMPLES OF JOB-CONCEPT FANTASIES

The following sections give examples of expressions of needs as they appeared in the Job-Concept interviews. Differences noted between the Disturbed and Comparison Groups are stated. Where the quotation reflects an attitude that was found as much in one group as the other, the boy is not characterized as to group. Certain constellations of needs were found to occur in patterns; these, too, are described. The needs are presented here in the order of Murray's list as modified for this study (see Appendix A).

1. *Acquisition*

Money, automobiles, homes, and other possessions were sought in the job fantasies. The better-adjusted boys contented themselves with reasonably high incomes, whereas the disturbed boys "waded through corridors of money." Usually another need underlies the Acquisition need. One boy, for example, wanted money so that he could placate his parents, while another boy felt that money would help him to dominate his friends. The first boy chose truck driver, the second elected commercial artist.

2. *Conservance and Order*

There was little difference between the disturbed and the better-adjusted in expression of these needs. They are associated with compulsive personality traits, and we found expressions of insistence on cleanliness, neat arrangement, and good working order. The better-adjusted boys more often abstracted these needs: an auto mechanic "helps the people of the United States to keep their cars going." A disturbed boy said, "It gives me a thrill to see things neat and working properly."

3. *Retention*

Retention fantasies were most often obtained from the boys in both groups who had difficulty in choosing a specific field of work. A disturbed boy's reluctant first choice was inventor. His other choices were mechanical engineer, radio operator, chemical engineer, and machinist. He said:

> If I couldn't have the others—if I was an inventor I'd be the others anyway—I'd be a machinist and chemist and anything else.

In analytic psychology, retention is associated with anality, with a complicated background that involves orality and relationship to the mother figure. The typical day of a disturbed boy who wanted to be an actor illustrates these connections:

> I have a big breakfast. I would have a big woman beside me. I like them real stacked. She has to do everything for me—dress me, undress me . . . I go to music places and cafeterias. The women want my money but I am a miser.

The opposites of Order and Retention appeared in the job fantasies too, and were rated as Outward Pain or Impulse Release. One boy wanted to be a roofer because he liked to smear tar about. His second choice was seaman, which he finally specified as mess-boy, because "I don't mind dirt—I like to get all greasy and dirty."

4. *Achievement and Recognition*

These needs were more often expressed by the better-adjusted boys. They related directly to pleasure in the work done and the wish to be recognized as skillful or to achieve status.

5. *Exhibition*

In both groups, this need was found to be associated not only with being seen, but also with the feeling of largeness. As might be expected, the boys who chose jobs associated with the stage showed this need. It also appeared in the choices of department store buyer, preacher, policeman, florist, surgeon, and sailor.

The subjects who chose acting and dancing emphasized that these occupations would give them the opportunity for self-expression. This is more in accord with the need for these boys to be admired for their persons than with professional theatrical theory.

The expressions of this need were almost stereotyped, regardless of occupation chosen: "I'm a big name in lights and people come to see me."

The emphasis on bigness was usually accompanied by expressions of inadequacy or the feeling of smallness:

> The universe is so big and people are so small. When I get to be an engineer, I'll be president of a gigantic firm and have a Cadillac a block long. My employees will greet me when I come to work in a silk top hat.

6. *Avoidance, Justification, and Attack*

Differences between the two groups were marked in the expression of these needs, which are associated with reactions to humiliation. The disturbed boys more often wanted to run away from humiliation whereas the better-adjusted boys were more likely to stand and fight off the humiliation through striving. This comes out clearly in these statements about math courses:

> Disturbed boy: Well, I didn't like the stuff so I just stayed away from class.
>
> Better-adjusted boy: I disliked math and I knew I'd flunk it if I wasn't careful so I paid special attention to it.

Avoidance was often linked with Inward Pain, and secondarily with other needs. One boy selected accounting, identifying himself with people who make errors, and avoided situations in which this might occur. He saw accounting as concrete, definite, and checkable. He expressed strong feelings of Achievement in his ability to work accurately with numbers:

> I like to work with numbers. I do it well—enjoy it really. . . . It comes naturally. There's an accomplishment you can see—a definite product produced. I feel competent when doing it. It's different when I do drawing—I feel uneasy—there might be mistakes.

While reality was often avoided, fantasy was not. A boy abandoned his wish to become a doctor because he believed it was too difficult for him. He chose refrigerator mechanic instead but clung to the fantasy of himself as a doctor:

> A refrigerator mechanic has to be like a doctor. He has to know what he is working with, which symptoms he is dealing with and what is wrong with the machine . . . dealing with something that has ceased to function and I put new life into it.

Justification, the offering of extenuation, was seldom encountered in either group.

7. *Dominance and Deference*

The following quotation from the interview with a boy who wanted to be a mechanical engineer illustrates our observation that the needs for Dominance and Deference were frequently found operating simultaneously:

> It's good to have someone over you and someone under you. If you want to be somebody you have to be able to tell people what to do. At the same time, you have to have someone to tell you what to do.

We were not surprised to find that Deference was associated with a passive role. The typical day of a draftsman was seen by one boy in this way:

> He is given ideas by his boss. . . . The boss tells him what he wants exactly and he does it. He finds out what the boss wants him to do and then he does it.

The better-adjusted boys more often offered rationalizations for Deference, stating that there was something to be learned from other people.

8. *Emulation*

Identifications were found with either the father or mother, or with both. A boy often identified not with a total parental figure but with an isolated trait he wished to incorporate into his own personality. The motive for the identification was also a complicated variable. Identification did not always mean that the boy wished to be like a loved or admired person; often the wish took another direction—to compete with and excel.

A boy chose to be a bookkeeper. He expressed a troubled leaning toward feminine activities:

> I like to cook but only for myself, not for anybody else. . . . I can't stand to see dust around. . . . I used to worry it (bookkeeper) was a girl's job—now I don't. . . . My brother and sister are no good at figures but me and my mother are.

Some of the identifications with the father were very direct. A boy said of his choice:

> My father worked with radios when he came over to this country. He would approve very much. Maybe he would have liked to be a radio technician himself.

Identification with the father sometimes was accompanied by anxiety, and the boy would mask his wish. A disturbed boy named baker (his father) as an aversion because "it's too hard work for me." His choice was restaurant manager (his paternal uncle) but his fantasy of the manager's work made him a cook and a baker. His made-up dream was as follows:

> I am in my uncle's shoes. I am running the place. I am doing what he is doing. I believe I would be working in the kitchen.

A better-adjusted boy chose mathematics teacher because:

> . . . the work is steady. My father has a good job but it isn't steady. I hope I can hold my job.

The disturbed boys more frequently identified with a trait or a feeling associated with another person:

> A friend of mine works (as a clerk) in the Chrysler Building. He says it's nice and seems to enjoy it. I'd like to be like him—he's friendly and knows how to act around people. In my character I'd like to have the good points in other people. Some people are too aggressive, others are too timid. I want to be even tempered. I need to be more aggressive. I've always been too shy. I'd like to be more polite, have more perseverance. I give up too easy.

9. *Autonomy*

Rebellion and independence are frequently assumed to characterize a phase of adolescent development. We found this true in nearly half of our entire population and slightly more often among the better-adjusted. The expressions were often the wish "to be my own boss," or "not to have to take orders."

Occasionally the need was more complicated. One boy resisted his mother's urging that he study upholstery and insisted, instead, on taking up electrical work—his dead father's occupation.

Another boy projected Autonomy into the work of chauffeur and linked with it the need for Tension Discharge, "You don't have to worry; you're on your own."

Autonomy was sometimes used as a defense against the boy's feelings of inadequacy and his dependency needs, linking the need closely to Dominance. The quotation which follows is from the record of a boy who wanted to be a shipping clerk, a job not likely to provide opportunity for Dominance:

> I don't want to be a salesman because you have to depend too much on other people. . . . I like this job because people depend on you . . . you can run it on your own. . . . I wouldn't worry what my friends think. I have to live my own life. . . . I would dream that I'm the boss. I come in and give orders. They all have to jump and I check on them. . . . People go for this kind of work mainly because they like to be independent and not always be told what to do.

10. *Uniqueness*

This need appeared slightly more often in better-adjusted boys. Their need was expressed in the wish to be different, to rise above the ordinary. When it appeared in the disturbed boy it was more often involved with the bizarre, or with the power of an engine or motor.

11. *Outward and Inward Pain*

Expressions of hostile and sadistic needs were disguised by the better-adjusted boy and more openly expressed by the disturbed boy. The need emerged in the concepts of a wide range of jobs: florist, boxer, teacher, minister, military policeman, scientist, and others.

Hostility with Justification is found in this statement by a boy who wanted to be a detective: "You have to be quick with a gun and not afraid to shoot to kill. After all, you have to protect citizens." A more open and joyous fantasy emerges in this concept of a military policeman: "I'd like to pick up a rat going AWOL—I'd shoot him dead."

Very striking was the greater frequency with which disturbed boys turned against themselves the aggression that earlier they had directed against others. Thus the boy who wanted to be a military policeman shoots a sailor in his made-up dream and then himself suffers a fractured skull.

Anna Freud (3) has described the child's tendency to "restrict" his ego as a defense against pain from external sources. She describes this defense as avoidance of activities (e.g., competitive ones) that are likely to produce pain and as a "normal stage in the development of the ego." We believe in a number of instances that the continuation of ego restrictions into adolescence is related to a masochistic tendency. One boy stated his wish to be an accountant:

> I'd like to make $25 a day; also my teacher in bookkeeping bawled me out and told me I really could become an accountant if I worked hard enough. (Then he restricts himself) I would mostly be like a bookkeeper. . . . I would worry at the end of the day that I had made a mistake . . . it's something that requires a lot of worry and responsibility.

Many boys were conscious of their ego restriction and expressed the feeling that they were not aiming high enough to win the peak of recognition. More than 60 per cent of all the boys spontaneously devalued their occupational choice (and themselves) in comparison with the physician: "People would hold it (physicist) near but not quite up to the elevation of a doctor."

When rarely a boy made a choice with an openly masochistic intent, it was always a dangerous occupation in which he foresaw injury to himself. Self-depreciation was seldom a direct determinant of the choice but rather was added to the vocation as a coloring:

> I'm in the printing plant. I start the machine and something goes wrong. I open the lock, check everything and don't find anything wrong. Something goes wrong again. I open it again and check it again—oil the press—I've thought of everything. Start it—something goes wrong. I stop it again. It's coming out blurred. So I fill up the ink, check everything. It runs o.k. The bad proof gets mixed with the good stuff and the boss gives me hell. I tell him I'll do it over at my own expense.

12. *Release and Restraint*

These needs are associated with the handling of asocial impulses, and their expressions reflect varying degrees of superego control. Expressions of masturbatory fantasies are included in this section because of the superego involvement and because such fantasies and practices are a concern of adolescent boys.

The expression of Release in the Disturbed boys was direct and frequently brutal. Thus a boy who feels the need to say, "I have good control of myself" goes on to say, "I'd write (as a journalist) about an animal that drowned . . . the most exciting thing about my day would be the fire . . . I wipe out the enemy." The better-adjusted boys showed more control: "A boxer has to fight hard. You want to kill the other guy but you have to obey the rules. Rules are important."

The hands were frequently involved in these fantasies. With Release these were expressions of pleasure in the use of the hands, in the joy of playing, manipulating, writing, handling, and in doing "dirty" work: ". . . painting, swabbing the decks, all sorts of menial tasks but I enjoy them . . . cleaning out the bilges."

Restraint expressions were accompanied by a shunning of manual work and a valuing of mental work, suggesting a displacement of affect from the hands to the head. A boy who wants to become a psychologist shows such displacement:

> When I have to use my hands I don't have quite the confidence I have when I can use my mind. Mental work is exciting. . . . My family would suspect, I mean, respect me. . . . I don't know whether I'll marry a girl who makes me feel good or who I can talk to . . . training is important because it overcomes human inadequacies. . . . I see no beauty in the body. If illness is psychosomatic you have to know the mind, not the body.

13. *Affiliation and Removal*

Very clear statistical differences were observed between the two groups in their expressions of these needs: the better-adjusted boys consistently showed more need for Affiliation, while the disturbed boys consistently showed more need for Removal.

Qualitatively, there were differences in the expressions of Affiliation by the two groups. The better-adjusted boys showed more of a sympathetic and of an independent attitude to the people around them in their work fantasies, while the disturbed boys were more dependent and more seeking of protection. The better-adjusted boys see themselves meeting girls on the job whom they marry and with whom they rear families. The disturbed boys express need for people around them to give them approval.

Removal, characteristic of more of the disturbed boys, usually was projected into the work concepts without explanation but sometimes with an explanatory note. This boy stated his reasons for wanting to be a policeman in a matter-of-fact way: "They go into it because you can walk by yourself and deal with people only casually. You don't have to come close to them."

Other boys attempt to overcome the need to withdraw by selecting an occupation that will force them into contact with people:

> You've got to sell yourself and that's what you've got to do. I'd like to get to know people better. It would help me get on with them better.

14. *Sympathy and Dependence*

The better-adjusted boys were consistently more sympathetic. The emphasis on Dependence shifted from the disturbed group in the first choice to the better-adjusted group in the second choice.

The expressions of Sympathy were varied, as were the apparent motives for the sympathetic attitude. The better-adjusted boys often ascribed an altruistic motive to their vocational interest: "I would design electrical equipment that would help people and make things easier for them." Variations in the motives for the sympathetic attitude emerged from statements like this one: "If I treat them good (his associates) they'll treat me good." This boy became the giver in order to become the receiver.

The sympathetic attitude sometimes is also a placating one serving to ward off aggression. A boy elected social work, stating, "I don't like people to dislike me. . . . If I found somebody I couldn't help it would hurt me."

Sometimes the sympathy need may be a mask for hostility, as with the child who starts to strike but converts the movement into a pat. Thus a boy who wished to become a florist mentioned the joy that flowers bring into people's lives, ". . . especially at funerals . . . there is so much death."

A disturbed boy saw the helpful social worker serving penance for wrongdoing: "Maybe they did something wrong in their lives and want to pay back by doing this work."

The frequency with which the need to be sympathetic was linked to the Dominance need suggested that a sympathetic, nurturing attitude is related to the need for an active role. This assumption finds support in the frequency with which the need for Dependence was found linked with a passive role. More than one-third of the boys injected a note about food or eating into their job fantasies, "A chauffeur would get in good with the cook. . . . Your boss tells you what to do."

The struggle between the wishes for active and passive roles sometimes was seen in the same fantasy:

> A teacher doesn't really think about his power too much. One teacher I knew realized his power, knew he influenced children and did his best to help them. . . . It's the same rut, but I enjoy it and feel secure. . . . It's a con-

> tinuous season, security, know how you stand, get regular vacation.

Sympathy is often expressed in a fantasy of restitution—giving back something that has been taken away. We have already cited one case in the fantasy about a forest fire[4] followed by replanting. In the restitution fantasy Sympathy is a denial of or a repayment for the boy's own aggressive fantasies. The same mechanism seems to be operating in the social worker fantasy[5] in which the person is making repayment for a wrongdoing. One more example from another field of work will show how pervasive we found the restitution fantasy to be:

> I have a television set to fix. I take the truck out and bring the set back. I bang it and it's all busted up. We overhaul the whole thing. . . . I'm happy and contented because I know what I'm doing and the people will enjoy the set.

15. *Tension Discharge and Accumulation*

The Discharge need appeared more often in the fantasies of the disturbed boys, whereas the better-adjusted boys showed more Accumulation need.

By our definition these needs express themselves in the relieving of tension by physical acting out of the tension or avoiding the acting out in favor of brooding or restraint. Discharge and Accumulation, therefore, are closely related to Impulse Release and Restraint.

16. *Asking and Telling*

The element of mystery and the seeking of solutions are intermingled with the expressions of Asking:

> Studying (as a doctor) would be important to me. There's so much I want to know. . . . The main thing is to find the source of illness. . . . You watch people come back from the dead to the living.

Many of the expressions of Telling involved the phrase "standing up." Thus ministers, teachers, and judges saw themselves "standing up" and telling people something, lecturing to or reprimanding them.

In several fantasies both the Asking and Telling need appeared (as in psychopathology, voyeurism and exhibitionism are frequently associated). A boy saw himself as a machinist:

> . . . listening to the other fellows telling how they would do the job. It's important to know what the other guy thinks or would do. Then I'd stand up and tell them how I'd do it.

V. CONCLUSIONS, IMPLICATIONS, AND SUMMARY

A. Conclusions

The salient findings are these:

1. The vocational choices of better-adjusted boys are more realistic than those of disturbed boys.
2. The second vocational choice of better-adjusted boys is *less* realistic than their first choice; whereas the second choice of disturbed boys tends to be *more* realistic than their first choice.
3. The vocational choices of better-adjusted boys are characterized by fantasies that emphasize participation in and involvement with the environment, and its inhabitants and objects. The vocational choices of disturbed boys are characterized by fantasies that emphasize removal from others and the environment, self-depreciation, and the acting out of impulses.

These findings support our hypothesis that individuals with different ego strengths will show differences in the role played by reality and fantasy in the making of their vocational choices.

Other findings suggest, but do not show conclusively, that adequacy of ego function may be related to:

1. Consistency in vocational choices—the making of similar vocational choices.
2. Level of functioning intelligence—measurements of intelligence by tests.
3. Homogeneity of functioning intel-

---

[4] See IV, A, 2*e*.

[5] See IV, B, 14.

ligence—consistency of performance in the various functions measured by tests.

4. Recognition of realistic opportunities for need-satisfactions—expecting similar satisfactions in similar situations and different satisfactions in different situations.

B. IMPLICATIONS

The results of this study are related to some general problems in psychology.

1. *A Theory of Vocational Choice*

No effort is made here to derive a total theory of vocational choice; rather the findings are related to extant theories and a contribution is advanced to the development of a theory.

The partial theory derived by Ginzberg (5) is that vocational choice is a developmental process in which the individual goes through a series of distinctive states progressing from choices based largely on fantasy to choices based largely on reality. Inherent in this progression is the growing capacity of the individual to make compromises between his fantasies and the realities of his world.

We found no evidence of a developmental progression towards greater realism of vocational choice, but much evidence that reality factors and fantasy drives operate simultaneously at all ages in the selection of a vocational goal. Our findings support only the "compromise" aspect of Ginzberg's theory. How is compromise achieved?

Compromise implies that reality is perceived by the individual, who uses these perceptions to curb, mitigate, or redirect the impulses generating fantasies. The perception of reality and the control of impulses are ego functions; compromise is a function of the adequacy or strength of the ego. We find evidence for this point in the greater realism of choices made by the better-adjusted boys, and the environment-involvement nature of their job fantasies as contrasted with the environment-avoidance fantasies of the disturbed boys.

A theory of vocational choice determination must take account of ego strength. With this factor as nearly constant as possible rather than an unknown variable, it would be possible to test the relative importance of developmental factors, interests, aptitudes, family traditions and pressures, environmental opportunities, and the host of other factors recognized as individual determinants but not yet integrated into a sound theory. Such a theory must encompass what goes into the choice, and the process by which the factors in play are consolidated into a choice.

We were unable to test the suggestions in the literature that specific needs tend to compel individuals into specific occupations. This is the concern of another investigation we have begun.

2. *Implications for Vocational Counseling*

The role of ego strength in the making of vocational choices restates what counselors have long realized: *vocational counseling, emphasizing as it does realistic factors, is best able to benefit the realistically oriented individual.*

The drive for need-satisfaction in work, its relationship to reality, and the dynamic balancing of these forces were demonstrated in our finding that the weaker ego is not without strengths. This explains another frequent observation: Counseling often is successful with emotionally disturbed individuals and when successful contributes to their stability. Counseling has helped these individuals to utilize and enhance their real strengths and it has also enabled them to satisfy certain needs and possibly to sublimate others.

Some counseling failures are explained by the adherence to an unrealistic choice. The implication is that such individuals have relatively weak egos which are unable to control the acting out of impulses or the search for unrealistic satisfactions. *Strengthening such weak egos is the problem of the therapist, rather than the counselor.*

Our qualitative analysis of the Job-Concept interviews suggests the basis for some typical problems encountered in counseling. One problem is that of the individual who cannot make a choice. Either he has endless fields he would like, or he can find no one field of work that attracts him. Our interviews point to the presence of retentive fantasies in some such individuals; specification of a vocational choice seems to mean renunciation of other possible choices and hence a loss which they cannot permit themselves to suffer.

Many young people adhere to the choice made at an early age and progress in their fields with apparent satisfaction. Others find their fields only on an empirical basis, requiring actual job experiences before they are able to affirm a choice. This exploration requires activity—seeking jobs and moving from one job to another in the effort to find the right one. Typical problems develop among these young workers. Some are insufficiently active in seeking work and once employed, remain there even when the opportunities for development have been exhausted. Others move from job to job, not remaining in one long enough to acquire the learning that the job provides. Lack of job-hunting activity or job movement may arise in a person with strong passivity wishes, a fear of taking an active role, or a masochistic tendency. Hyperactivity in job movement may arise as a defense against deference wishes, having needs that remain unsatisfied by the job, or from a strong acting-out impulse.

Another frequently encountered set of problems is that of young people who overevaluate themselves in the making of their choices. They seek to enter vocations for which they have little preparation or promise. If their educational background is inadequate, they steadfastly refuse to complete or strengthen their qualifications. Overevaluation may arise in adolescents with a strong exhibitionistic need, associated with a wish for bigness, and a need to offer extenuations for inadequacy rather than to attack the problem by striving.

Those who underevaluate themselves make choices far less demanding than the intelligence, skill, or responsibility of which they are capable. These young people show the operation of a self-imposed ego restriction which is linked with masochism.

All of this points to the necessity for counselors to have the means for identifying and working with the personality needs of their adolescent clients. The means would include instruments for eliciting and evaluating needs (e.g., the Job-Concept interview) and the personal training and insight to relate such findings to the counseling of the individual. Counselors also need knowledge of the range of possible satisfactions or sublimations that are afforded by each job area. This knowledge is usually already available to the skilled counselor through his intimate acquaintance with varieties of jobs, their slight internal variations as well as marked ones, their training requirements, and their availability. The job knowledge which most counselors already have would achieve its greatest utilization when coupled with a method for identifying the pressing needs of a young person as he relates them to his vocational choice.

3. *Implications for Further Research in Ego Functions*

This investigation has established and measured one ego function—realism of vocational choice. The data also indicate that consistency of vocational choices, functional level, and homogeneity of intelligence, and recognition of realistic opportunity for need-satisfactions may be aspects of ego function. With further research and development of instruments, these, too, may be measurable.

What could be served by further efforts to measure ego functions and functioning? Education and psychotherapy will be considered here.

Teachers have long recognized the difficulties of the "average" program for the child of below- or above-average intelligence. More puzzling has been the bright child who learns poorly, or the dull child who learns well—each at a level of instruction geared to his intelligence. Inquiry into the ego functioning might illuminate this problem. Better programming and grouping of students might be the reward of information about the way in which ego strength affects learning when the IQ is constant.

Psychotherapists feel the need to evaluate the progress of an individual under treatment, and to compare the relative effectiveness of different treatments. Clinicians have tried "before and after" studies using projective testing methods, but without satisfactory results.

This state of research affairs may arise from the very nature of projective techniques and their stimulation of primary impulse processes which change very little, if at all, in the therapeutic process. What does change in successful therapy is the ability of the individual to mediate between these impulses and the demands of reality—the strength and functioning adequacy of the ego component of the personality.

The delineation of ego functions and the development of instruments for their measurement should further the ability to evaluate both the therapeutic progress of individuals and the effectiveness of different treatments.

C. SUMMARY

This study explored the interaction of reality and fantasy in the making of vocational choices. Assuming that vocational choice is an ego function, we selected two groups of subjects: 50 better-adjusted boys with relatively strong egos between the ages 15-19, and 50 disturbed boys of the same ages with relatively weak egos.

Vocational choices were obtained from the subjects, who were all receiving vocational counseling. The boys were rated by their counselors for the degree to which they possessed the personal characteristics required by the vocations they had selected. The extent of deviation of characteristics possessed from characteristics required was termed a Reality-Deviation Score. The better-adjusted boys were consistently more realistic in their choices than were the disturbed boys. The data also suggest that a positive relationship may exist between the selection of similar occupations for the first and second choices and (*a*) higher IQ, (*b*) less subtest scatter on the Wechsler-Bellevue, and (*c*) greater realism in relating personal characteristics to job requirements. It is suggested that these factors are to some extent ego functions.

Within the age group studied, no evidence was found supporting Ginzberg's theory that occupational choice is a developmental process leading chronologically to increasing realism as a determinant of choice.

Evidence is presented for accepting the reliability and validity of the Job-Concept interview method for assaying the fantasy or needs content of vocational choices. The three methods employed for assessing needs—the Job-Concept interview, the Case History, and the Thematic Apperception Test—vary in their sensitivity to the 27 needs rated, with the Job-Concept interview being sensitive to more needs than either of the other two methods.

The better-adjusted subjects show more of the needs that are associated with environment-involvement, forming relationships, and the exercise of skills and talents. The disturbed subjects show more of the needs that are associated with environment-avoidance, and the restriction of both relationships and exercise of talents. Disturbed subjects show more of a tendency to seek satisfaction of the same needs in each of two dissimilar vocations.

Some implications of the findings for a theory of vocational choice, for vocational counseling, and for further research in ego functions are discussed.

APPENDIX A

Definitions of needs used in this study are presented here. We adopted Murray's (9) needs, making a few changes and additions, simplifying the terminology and standardizing the definitions as follows:

A. NEEDS ASSOCIATED WITH INANIMATE OBJECTS

1. *Acquisition.* To gain possession of property. To grasp, steal, or snatch things. To bargain or gamble. To work for money or goods. To borrow.

2. *Conservance.* To collect, repair, clean, and preserve things. To protect against damage.

3. *Order.* To arrange, organize, put away things, be tidy and clean.

4. *Retention.* To retain possession of things. To refuse to give or lend. To hoard, be frugal, economical, and miserly.

B. NEEDS ASSOCIATED WITH AMBITION, ACCOMPLISHMENT, AND PRESTIGE

1. *Achievement.* To overcome obstacles, to exercise power. To do the difficult quickly and well. To emphasize the achievement.

2. *Recognition.* To excite praise, respect. To boast, seek distinction, social prestige, honors, high office. This need is related to the next.

3. *Exhibition.* To attract attention to one's person. To excite, amuse, shock, and thrill others. Self-dramatization.

C. NEEDS ASSOCIATED WITH HUMILIATION

1. *Avoidance.* To avoid failure, shame, humiliation, ridicule. To refrain from doing something beyond one's power. To conceal a disfigurement.

2. *Justification.* To defend self against blame or belittlement. To justify one's actions. To offer extenuations, explanations, and excuses. To resist probing.

3. *Attack.* Proudly to overcome defeat by striving and retaliation. To select the hardest tasks. To defend one's honor in action.

D. NEEDS ASSOCIATED WITH HUMAN POWER

1. *Dominance.* To influence or control others. To persuade, prohibit, dictate. To lead, direct, restrain. To organize behavior of a group.

2. *Deference.* To admire and willingly follow a superior. To serve gladly.

3. *Emulation.* To empathize, imitate, emulate, identify with. To agree and believe.

4. *Autonomy.* To resist influence or coercion. To defy authority and seek freedom. To strive for independence.

5. *Uniqueness.* To act differently from others. To be unique. To take the opposite side. To hold unconventional views. To be bizarre.

E. NEEDS ASSOCIATED WITH INJURY AND PAIN

1. *Outward.* To assault and injure. To belittle, harm, blame, accuse, ridicule. To punish severely.

2. *Inward.* To surrender, comply. To accept punishment. To apologize, confess, atone.

F. NEEDS ASSOCIATED WITH ASOCIAL IMPULSES

1. *Release.* To disregard social and conventional inhibitions. To act impulsively regardless of consequences, blame, or punishment.

2. *Restraint.* To avoid blame, ostracism, or punishment by inhibiting asocial or unconventional impulses. To be well behaved and obey the law. To inhibit the impulse.

G. NEEDS ASSOCIATED WITH AFFECTION BETWEEN PEOPLE

1. *Affiliation.* To form friendships and associations. To greet, join, live with others. To cooperate. To join groups.

2. *Removal.* To snub, ignore, exclude. To be aloof, indifferent, discriminating.

3. *Sympathy.* To nourish, aid, or protect. To be sympathetic, motherly.

4. *Dependence.* To seek aid, protection, sympathy. To cry for help, plead for mercy. To be dependent.

H. NEEDS ASSOCIATED WITH TENSION

1. *Discharge.* To relax, amuse self, laugh, joke, play games, avoid serious tension.

2. *Accumulate.* To brood, worry, mourn. To avoid or delay release of tension through physical activity or companionship.

I. NEEDS ASSOCIATED WITH INFORMATION AND KNOWLEDGE

1. *Asking.* To explore, question, satisfy curiosity, look, listen, inspect. To read, seek knowledge.

2. *Telling.* To demonstrate, to relate facts. To give information, explain, interpret, lecture.

REFERENCES

1. ANGYAL, A. *Foundation for a science of personality.* New York: The Commonwealth Fund, 1941.
2. FEATHER, D. B. The relation of personality maladjustments of 503 University of Michigan students to their occupational interests. *J. soc. Psychol.*, 1950, **32**, 71-78.
3. FREUD, ANNA. The ego and the mechanisms of defense. New York: International Univer. Press, 1946.
4. FRIEND, JEANNETTE. Personal communication, 1951.
5. GINZBERG, E., GINSBURG, S. W., AXELRAD, S., & HERMA, J. L. *Occupational choice.* New York: Columbia Univer. Press, 1951.
6. HART, H. Work as integration. *Med. Rec.*, 1947, **160** (December), 3-12.
7. HENDRICKS, I. Work and the pleasure principle. *Psychoanal. Quart.*, 1943, **12**, 311-329.
8. LEWIS, J. A. Kuder Preference Record and MMPI scores for two occupational groups. *J. consult. Psychol.*, 1947, **4**, 194-201.
9. MURRAY, H. A. *Explorations in personality.* New York: Oxford Univer. Press, 1938.
10. ROE, ANNE. Personality and vocation. *Trans. N. Y. Acad. Sci.*, 1947, **9**, 257-267.
11. SMALL, L. The role of self-esteem in the rehabilitation of emotionally disturbed and delinquent adolescent boys. Ph.D. Thesis, New York Univer., 1950.
12. WAR MANPOWER COMMISSION. *Dictionary of occupational titles, part iv.* U. S. Government Printing Office, 1944.

(Accepted for publication June 25, 1952)

LIBRARY
12.8.7
9158

Bureau Ednl. Psy. Research
DAVID HA [illegible] AINING COLLEGE
Dated ....
Acc. No ....

Vol. 67, No. 2 Whole No. 352, 1953

Psychological Monographs: General and Applied

A Psychological Study of Eminent Psychologists and Anthropologists, and a Comparison with Biological and Physical Scientists[1]

Anne Roe
New York, N.Y.

I. INTRODUCTION

THIS is the third and last monograph in a series of clinical studies of research scientists. The series of studies was designed to investigate the existence of relationships between life history, intellectual functions or personality characteristics, and the selection and pursuit of a particular science as a profession. This has been the first series of its kind in this field, and hence the major approach has had to be observational and diffuse. In so complex a problem, the first need is to get some idea of the nature of the relationships, it any exist, the points at which a direct attack can be made, and the sort of tools to use. It was felt that no existent personality theory was sufficiently developed, or generally suitable for the derivation of hypotheses in advance. Now that extensive observations have been made in this specific field, for this specific purpose, it is possible to set up a number of hypotheses concerning these relationships which can be checked directly in future work.

The subjects of the study are men who were selected for their eminence in research, as judged by their peers. The data comprise verbatim life histories, discussion of the work of the men, and results of three tests, a Verbal-Spatial-Mathematical Test, the Thematic Apperception Test, and the Rorschach. In addition there were obtained, for comparison with these groups of eminent men, group Rorschachs of members of university faculties in the same fields.

The rationale of this approach is discussed in the first monograph, which presented the results of the study of biologists (21). The second monograph (22) is a similar study of physical scientists. This monograph reports the results of the study of psychologists and anthropologists, and a comparison of all of the groups. A short paper analyzed the interrelations of the tests of individual biologists (17), and another analyzed test interrelations for the total group of scientists (20). A separate paper on the scientists' use of imagery has appeared (19), and the group Rorschach studies have been reported in full detail in four papers (16, 18, 23, 24).

It seems almost impossible to formulate an adequate expression of appreciation to the subjects who have served in all of these studies. Their gifts, not only of time and effort, but of personal revelation, have resulted in data unique in psychological annals. It is with profound realization of the extraordinary privilege it has been to gather these data, and with a deep sense of obligation that I have attempted to organize the material for others.

For obvious reasons the names of the subjects have been withheld in these publications, and although many of them are recognizable from their histories, the test data cannot be related to these. Present eminence is not a sure criterion of future eminence, but it is evident that the clinical psychological records of these men would be of extraordinary interest to biographers. With the consent of the subjects, arrangements have been made to leave the full accounts to the library of the American Philosophical Society, where they will be available to research workers in due course.

---

[1] This research was supported by a grant from the National Institute of Mental Health, of the U. S. Public Health Service. Publication of this monograph was made possible by a grant from the Wenner-Gren Foundation.

II. SELECTION OF THE SAMPLE

The sample restrictions imposed in the previous studies were also observed in this, so that the subjects are all men, under 61 years of age, American born, and currently engaged in active research. The preliminary list of 94 psychologists was constructed with the aid of Dr. E. G. Boring and Dr. David Shakow. A few other names were added by the raters (E .G. Boring, E. R. Hilgard, D. B. Lindsley, J. W. Macfarlane, D. Shakow, and L. M. Terman), and the men were ranked on the basis of the combined ratings. The raters were in close agreement for the experimental psychologists, but agreement was very poor for all of the others, with no close correspondence between fields of rater and subject. This would seem to reflect a general confusion over standards for research that is not strictly experimental in character. In the final group there are, in broad classification, 10 experimentalists (comparative, learning theory, sensory experimentation, etc.) and 4 whose major fields lie in clinical, social, and developmental psychology, and testing. Three experimentalists refused, one because of serious illness, and one of the others could not be included when illness made it impossible to arrange for the interviews.

The original list of anthropologists, constructed with the help of Dr. Robert Lowie, included 82 men. The raters were A. V. Kidder, A. L. Kroeber, R. H. Lowie, A. H. Schultz, and C. F. Voegelin. Of the 13 men selected on this basis, one was out of the country, 3 refused, 1 was ill, and the other 8 are included in the study. There are two physical anthropologists, two archeologists, and four cultural anthropologists.

Although the data for psychologists and anthropologists are presented separately, these groups are combined for comparison with previous studies and together referred to as the social scientists. The term is not very satisfactory, particularly for experimental psychologists, but is probably better than devising a new rubric.

Excluding from each group those who were unable to cooperate because of illness or geographical location, the percentages of refusals were: biologists, 13%; physicists, 27%; social scientists, 22%. The higher rate of refusals among the physical scientists is, I believe, in considerable measure due to their heavier schedules.

III. DESCRIPTION OF THE GROUP

The psychologists have an average age of 46.7, the anthropologists 49.4; for all social scientists the average age is 47.7. The physicists averaged 44.7 years and the biologists 51.2.

Family Backgrounds

Enough data were gathered on the families of each subject to give some idea of the general socioeconomic position. No two psychologists were born in the same state, but four came from the East, six from the Midwest, and three from the West. Four of the anthropologists were born in the East, three in the Midwest, and one in the West.

Occupation and education of the fathers of the subgroups are given in Table 1. Half of the psychologists and three of the anthropologists had professional fathers. Only 4 of the 14 psychologists came from families with rather

TABLE 1

OCCUPATIONS AND EDUCATION OF FATHERS

Occupation of Father	No. of Fathers of		Education of Father	No. of Fathers of	
	Psychologists	Anthropologists		Psychologists	Anthropologists
Physician	2	0	Elementary	4	1
Engineer	2	0	High school	0	3
College teaching	2	0	Some college	4	2
Lawyer	1	3	College graduate	5	2
Businessman:			Unknown	1	0
Own	2	4			
Clerk, agent	2	1			
Skilled labor	1	0			
Farmer	2	0			

good incomes, but 7 of the 8 anthropologists came from families that were well-to-do. Incidence of professional fathers in the other groups was 45% for the biologists and 73% for the physicists (experimentalists 50%; theorists 84%).

TABLE 2

NUMBER OF CHILDREN IN PARENTAL FAMILY AND BIRTH ORDER OF SUBJECTS

No. of Children, Including Subjects	No. of Subjects		Position in Family	No. of Subjects		
	Psych.	Anthro.		Psych.	Anthro.	All Scientists
1	3	3	1	6	5	39
2	1	2	2	3	2	13
3	5	2	3	2	0	3
4	4	0	4	3	0	3
5	0	0	5	0	0	2
6	1	1	6	0	1	2
			7	0	0	2
Average	3.0	2.4				
Average	2.8					

Number of children in the parental family and birth order of these subjects are given in Table 2. The distributions are similar to those for the biologists and physicists, and birth order for the total group of 64 scientists is also given. Comparison of the observed number of first-born with the calculated expected number shows the incidence of the first-born in these groups to be reliably greater than chance ($p < .01$).[2] Cattell and Brimhall (5) also remarked on the high incidence of first-born in their sample of 855 scientists. For their group it was possible to check incidence of first-born in each family size from 2 to 7 and in all of them it proved to be greater than chance.

Of the 25 scientists in my group who were not first-born, 5 are oldest sons, and 2 of the second-born were effectively the oldest during their childhoods because of the death of older sibs, one at birth, one at age 2. Complete data are not

---

[2] I am indebted to Dr. Howard Levene for these calculations. His method is as follows: For families of size s, and n cases, the number of first-born has a binomial distribution with mean $np = n\,(1/s)$ and variance $npq = n\,(1/s)\,(s - 1/s)$. When we add number of first-born for families of different size, the expected number is equal to the sum of expected numbers and since different families are independent, the variance of the sum is equal to the sum of the variances. The mean and the standard error of the total first-born will be approximately normally distributed if the total number of cases is large.

TABLE 3

AGE AT MARRIAGE AND NUMBER OF CHILDREN

Age at Marriage	No. of Subjects	
	Psych.	Anthro.
33	1	0
32	1	0
31	0	1
30	1	0
29	0	0
28	1	1
27	3	1
26	2	1
25	1	1
24	1	0
23	2	2
22	0	0
21	0	1
20	1	0
Average...........	26.5	26.1

No. of Children	No. of Subjects	
	Psych.	Anthro.
4	1	1
3	1	1
2	6	2
1	4	2
0	2	2
Average..........	1.6	1.6
Sons.............	11	8
Daughters........	12	5

available for 3, but for the others the average number of years between the subject and his next older brother was 5. The possible significance of this will be discussed later.

MARITAL STATUS

All of these men are married and most of them have children. Data are given in Table 3. Average age at marriage of all three groups of scientists is rather late, which is doubtless in part connected with the long educational histories. The social scientists, however, differ greatly from the others in the permanence of their marriages. Among the biologists, there have been three divorces (15%); among the physicists, one (5%); but five of the psychologists (36%) and four of the anthropologists (50%) have been divorced, and of these several have been divorced more than once.

That the psychologists and anthropologists have a smaller average number of children than the other scientists may be related to the difficult marital histories, but the differences are slight. The sex ratios of the children are quite different in the different groups of scientists but this is probably chance: daughters are much more numerous for the physicists, sons for the biologists; for the social scientists, the difference is slight.

COLLEGE AND GRADUATE SCHOOL HISTORIES

Ages at which these subjects completed various stages of their formal training are given in Table 4. The averages are

TABLE 4

AGE AT RECEIVING COLLEGE DEGREES

Age at Receiving	B.A. or B.S.		Earned Sc.D. or Ph.D.	
	Psych.	Anthro.	Psych.	Anthro.
32	0	0	0	1
31	0	0	0	3
30	0	0	1	0
29	0	0	2	1
28	0	0	0	0
27	0	0	2	1
26	0	0	2	0
25	1	0	1	0
24	1	1	5	2
23	1	2	1	0
22	1	2	0	0
21	6	3	0	0
20	4	0	0	0
Average age...	21.4	22.1	25.8	28.6

Psychologists: 9 took M.A., average age 23.8 years.

Anthropologists: 7 took M.A., average age 24.4 years.

about the same as those for the biologists, but about a year higher than those for the physicists. More of this group than of the others took Master's degrees.

Five psychologists and four anthropologists received their Ph.D.'s from Harvard, two psychologists from Yale, and one of each from Chicago. The others all attended different institutions.

PROFESSIONAL HISTORY

Since completing their formal schooling the members of this group have moved about relatively little. The number of institutions with which they have been connected professionally is given in Table 5. The distributions and means are similar to those of the earlier groups, with the slightly higher mean for the biologists probably reflecting their higher average age.

TABLE 5

NUMBER OF INSTITUTIONS WITH WHICH SUBJECT HAS BEEN CONNECTED SINCE DOCTORATE

No. of Institutions	No. of Subjects		
	Psych.	Anthro.	Both
4	2	2	4
3	4	0	4
2	5	2	7
1	3	4	7
Av. no. of institutions.......	2.4	2.0	2.2

Nine of the social scientists had postdoctoral fellowships—five National Research Council, one Social Science Research Council, one Guggenheim, and two others. Most of the anthropologists have had grants for field work.

Among the psychologists there have been no major changes in field of work although several have shifted gradually until they are now rather far from where they started. There have been greater changes among the anthropologists. These seem to have been due to contact with particular persons, or to the specific nature of the jobs which were available. Nevertheless, the general pattern is of being able to control to a fairly considerable extent the nature of the research undertaken.

IV. EARLY HISTORY RELEVANT TO OCCUPATIONAL CHOICE

One of the striking differences between the social scientists and the others is the amount of material which was spontaneously offered in the interview. In part this may be due to the greater understanding among this group of the general problem and the relevance of details of personal history, but in perhaps larger part it is a reflection of their greater ease of verbalization. This means, among other things, that the life histories for this group must be subjected to considerably more cutting than was necessary for the earlier groups.

The interviews were very little structured. The subject was asked at the outset for information on general family background, early family and school life, and everything he could remember that related to his choice of vocation. I interrupted as little as possible, usually only to clarify a point or to recall him to pertinent material. Later I asked specifically for information on health, religion, and present leisure interests, as well as on use of imagery. Sometimes the projective material suggested questions for later intervals, but under these circumstances deep probing is impossible. The combination of projective material and life histories recorded verbatim offers excellent cross checks.

The histories which follow are first discussed separately for the psychologists and the anthropologists. They are arranged in a chance sequence which bears no relation to the code numbers assigned to each man for presentation of the test data.

The Psychologists

A. "My father was the ideal of a country doctor and he took me around with him on visits to his patients. He was on the earthy side. Mother was brought up in a highly charged religious culture but she broadened later.

"In grade school I wasn't a distinguished student. I didn't like school but it wasn't a particular problem. I was just doing what I was supposed to do. I suffered agonies on the playground because I couldn't chin myself and things like that; I think it was more my constitutional type than ill health. I started something that has been a life pattern, to find out what I could do and do it well. I was very clever at sidestepping from childhood on. I surrounded myself with those I could do business with and I avoided others. I have always avoided fights and competition but I always had playmates. I wasn't elected to class office but I did become editor of the school paper.

"In high school I had a girl friend and was a rather happy adolescent provided I could keep out of competition. I made my own world, I always have. The odd thing about me was I had no aspirations at that time. I felt driven to do the damn best I could but I don't think it was because of aspiration or ambition or egotism, but I was haunted, driven. Even now I can't take a vacation without taking work with me.

"I went to college largely because of a brother who was there. My college experience was just a new world. There wasn't any world until then hardly for me. I just took things for granted. Intellectually, college was a marvelous opportunity, but this was not only intellectual. Another element that played a very large part was the experience in social service. I had a boys' club and became much interested in social work. It seemed to release in me an idealism I hadn't been aware of before. I got a tremendous kick out of doing good and that interest always was a close second to my studies. I studied hard, first because I always felt driven and partly also because Father had given me to understand that I would have to have a scholarship every year, and I did. If I had vocational ideas through college they were in the direction of social work rather than science. I didn't know what I was going to do. It was funny that I wasn't worried about the future.

"I taught abroad for a year but never thought of anything beyond that. Then the family insisted that I come home and I got a graduate scholarship. My undergraduate major was called economics but I think I went into psychology to find out what one of my professors was talking about. He intrigued me. There was no doubt that I had an authentic learning experience, although he was totally unintelligible to me. In two years I had my Ph.D. I was pretty immature and the department was really very weak as I see it now. I went abroad again on a fellowship. It took me some months to find out what was going on. But I got oriented a bit and I began to learn what to read in the new German psychology to which I was exposed. Having to make my own way (intellectually) meant a lot. I don't like guidance. I don't like spoon-feeding. I don't like pampering. It's much better to be thrown on your own."

B. "My father was the son of a businessman in upper New York State. He didn't have to earn a living or do any serious work until he was over thirty. He had studied voice abroad and then he took a job as singing teacher in a small college. My mother went to a ladies' seminary and taught dramatics in public school.

"I went from kindergarten and through junior high school in a laboratory school associated with the college and I did pretty well all the way through there. Then I went down to the public schools for the last two years of high school and my marks fell apart. The teaching was mediocre and I was less at home. The time coincided with the development of heterosexual interests and I was much more interested in going out with the girls than I was in studying. I had no intellectual interests. As a boy the thing I liked was being out in the woods. I was interested in reading. I read a lot but it was mostly Edgar Rice Burroughs and that sort of thing.

"I started in a teachers' college because we were living right there. I didn't have any ambitions. For a while I was interested in journalism. It was just taken for granted that I would go to college. Father had graduated and mother had the equivalent of a college education. But after the first two years I said to hell with it. I'm not going on with college. It was just a source of tension and stress to me. I went to work on a railroad construction gang and I damn near killed myself to prove that I was a man. By fall I decided that earning a living with your back muscles wasn't so good, and I went back to normal school. That's when my academic record improved. I took an English major, probably because of deficiencies. I disliked math and that threw out the sciences immediately. I wasn't inclined toward art and music was out because of the family situation. I guess I had secondary school teaching in mind. I seemed to have no alternative ideas, and it was a way to earn a living. I enjoyed my teacher training and thought I did it well. They gave us a chance to work out our own plans and I really enjoyed that.

"Then this is the way I got interested in psychology. A fellow who had just gotten his Ph.D. came into the psychology department, and I first saw the light as far as my personal problem was concerned. I realized that I wasn't deviate and that there was some hope for me. Then he became a father substitute. He let me go to his home and would talk to me on an equal level and he got me very excited about scientific objectivity. I got my B.A. in a depression year and there were practically no jobs. At the same time they had just started giving an M.A. and he offered me a fellowship which was enough to pay Mother for board and room and keep up my self-respect. So I stayed and got an M.A. and I was more and more interested. I had been reading Lashley, some of which I still don't understand and much of which I didn't understand then. I did an animal study all by myself and no one bossed me around or even helped me. By that time I decided I wanted to go into professional psychology.

"I taught high school for a year and then went back with an assistantship at another university. I did a very hard job on my thesis. I literally lived in the laboratory for the better part of the year. The professor didn't give me any encouragement but indicated indirectly that he had confidence in me. He was just the sort of guy who would leave you alone but you could have contact if you were persistent enough about it. And he was an eminent man and that made a lot of difference to me."

C. "Father was a construction engineer and finally became very prosperous. He had a good mind and a college education but he was most certainly not an intellectual, and tended on the whole to scorn professors and such. Mother went to college but I can't remember if she finished. Any expression of her intellectual side has always been very much restricted by her strong primitive religious background.

"I learned to read long before I went to school, but I have really no idea how. It's characteristic of the family situation that the first book I read was a fat Bible story book. Reading was pretty nearly all my life in the early school days, and writing. I liked all types of school work; looking back on it I liked them too much. I was a shy youngster. I can't believe I had any social adjustment in the group. I had enough companions at home to make it unnecessary. In our home, reading was only sort of all right; if you had done your chores and there was nothing else to do it was okay to read. About the time I was ready to start high school my parents bought a farm and I had to transfer several times (because of transportation difficulties). I was practically always the top of my grade in most everything. My social life was unbelievably nil. Dancing, card playing, and movies were verboten, and even carbonated beverages weren't quite right.

"There was never any question about going to college. We went to the state university where father had gone and started living at the Y dormitory. I knew what I wanted to do. I'd always been interested in nature. It was about the time of Gene Stratton-Porter and her books about moths of the Limberlost figured very much. I started reading about moths and finding caterpillars and feeding them and I got to be quite an authority on night flying things. Then father wanted the farm run scientifically and he had some of the agricultural people come down from the state university. That began to fascinate me. I realize now that what intrigued me was the scientific part. Agricultural work is good experimental work and I took that in. So I registered in agriculture and had every idea that what I was going to be was a farmer.

"Being of a somewhat emotional and sensitive sort, religious conferences at the Y hit me pretty deep and I don't know whether gradually or suddenly I began thinking of the ministry. I realized a broad background would be a good thing and I changed my major to history. And then, too, the exploring, the scientific, the scholarly side appealed to me more and more.

"Then I was able to attend a conference abroad. That was an education, my gosh. That stretched my thinking enormously and pretty well disabused me of a lot of the narrow religious ideas I had. It determined me to go to some liberal seminary and it strengthened my decision to go into the ministry.

"The motivation was really a service motivation. Quite a strong one, I should say, and on a fairly abstract sort of level in a sense. I found the seminary an extremely stimulating place. They really believed in freedom of thought and inquiry and there was a bright group of students and in no time at all we were teaching ourselves. In courses in religious education we got a good deal of what really would be clinical psychology. That appealed to me very much and sort of shifted my focus to religious education. On the intellectual and philosophical side there was a steady growth in questioning on the part of the whole group of us. I began to take courses in psychology and by the end of my second year I definitely decided on it. I was interested in child guidance work and the service motivation was definitely dominant.

"I changed to educational psychology and I got a fellowship in child guidance. It was a fairly tough year. At the university they were rigidly objective, emotions didn't count and the Child Guidance Institute was everything from ultra-Freudian to statistical. That was very fruitful, that was awfully good training. That year I began to realize I have a facility for working with people. Then I went to another social agency that gave me a kind of chance I think not enough people get. I just got a snoot full of work. I wasn't particularly thinking about what to do next professionally, there was just so much work to do, so many children to see, so many agencies wanting help with children that I just got deeply immersed in the clinical function. I just learned how to work with kids. And there was none of what I feel has so often killed clinical psychologists, there was no one whom we had to be subservient to."

D. His parents came from abroad but were married here. His father made a rapid ascent in business, but died when the subject was in his early teens. He went to the public schools but had a good deal of illness and was left back.

"I went through school doing competently, never being very much engaged. It was never a really challenging kind of thing to do. I spent a lot of time in the woods and I did a lot of reading of very conventional things. After father's death we moved about a good deal. It was very confusing for the next few years. I went to about six different high schools until I was finished and there was always the problem of establishing myself. The critical problem of staying and getting a footing became more and more important to me. I began to develop some special ways of doing this.

"Then I ran on to the first teacher who really interested me, in a course in European history. I started reading in the field and I read outside of school and I became interested in working at maps. That course was the first real satisfaction I ever got in a course. It represented material which could be organized, you could make sense out of complexity. You could see elegant trends.

"I wanted to go to college. It was taken for granted that I would go and it was somehow assumed that I would take a pre-law course. I did for two years and I was very indifferent and uninterested. Meantime I was putting in a lot of time at writing. Not stories, but essays, trying to get my thoughts straight. Writing a lot to the college newspaper, for example, and one thing and another like that.

"At the end of my second year I took a course in psychology and the tremendous comprehension and dogma intrigued me and infuriated me. I was intrigued by the fact that he had tried to cover so much in one simple theory. I started thinking and I started reading. Then I suddenly realized that this was the stuff that interested me and in a curious kind of way, and the term after that I decided to go ahead and find out more about it. I couldn't do anything by myself so I took some more courses. I felt a certain impotence.

"That was a period when the fever started developing. One of the professors took a group of us and he thought if we wanted to learn about things, the way to do it was to do research. My senior year I carried through some research. That really sent me. That was the thing that trapped me. After that there was no getting out. I tried anthropology and it infuriated me. I enjoyed it thoroughly, but I found that it just didn't satisfy me. That insofar as one could design anything elegant, it would be banal. At the same time all these things burst up together. I began to have some sense of what constitutes poetry. It was another world just opening up before me, and the same thing happened in music. All this in a period of a half year. I remember I decided that I wouldn't go home for Christmas that year. It was a kind of painful thing but I just could not communicate with the family so I didn't go.

"By this time I knew for sure that I wanted to be in psychology but I didn't know what I wanted to be in psychology. I did some research projects and it seemed to me all the wonderful fiery dogmas of the period were all wrong. That first year of graduate work was heavenly. Then I developed a new, more specific kind of promiscuity. I spent a summer working on operative techniques after devouring Lashley's papers. I read enormously in physiology and in anthropology and I started working on field problems with monkeys. I didn't finish because that spring I somehow had the feeling that that was getting to be too much of the same sort of thing and that the field of physiology was not quite all I wanted. I read Allport's *Personality* and Boring's *History* and I decided by gosh I'd go on."

E. "I was born in a town of about 5,000 and we lived under very restricted financial circumstances. I wandered through grade school, but in high school I was the top of the boys. I expected to become a teacher. Socially I was very restricted. My father was opposed to our learning to dance. I didn't date in high school although I would have liked to. I was small and the group I went around with was younger, I found I could play football with them. I was quiet and never caused any trouble. I did my lessons, my teachers liked me. My oldest brother thought I was too mild so he taught me to box. My brothers were very close. Even in college my closest friend was my next older brother. We lived close enough to poverty, and always in the family there was a great confidence in our complete intellectual superiority. There was a very strong intellectual emphasis on this kind of thing.

"I would say I did a moderate amount of reading. I wasn't a voracious reader. I didn't have any particular intellectual interests in high school. I think college was assumed, and it was taken for granted I would go to the one in my home town. Then an older brother went away to college and my aunts financed my going. The one thing that surprised me was that I wasn't homesick. I had expected to be. There I developed my first real intellectual interests. I had no concept of studying. It wasn't until I reached my second year that I found people studied for exams and it was a shock. The first thing I found I liked was biology but I was revolted by dissection. Undoubtedly I was just drifting. I gave my major as English because my brother did. I liked history very much—I liked memorizing dates, facts, people, when I found out you were supposed to do it. When I took psychology I pretty well decided to become a psychologist. It was interesting like biology and as far as I could see you didn't have to dissect. I was as interested in the physical sciences as in the psychological but I had an emotional block against the mechanical aspects. The chairman of the department thought I had the ability to go into graduate work and the last year I was given a teaching assistanship; I had had a job running rats on an hourly basis. I avoided advanced work in statistics, let's be honest, I was scared of it. I don't think I was unusually inept but I lacked the blind confidence."

F. "As far back as I know both my parents' families were farmers with no professional people among close relatives. My parents had only a country school education. Father was quite an intelligent man and did quite a bit of reading. My home was favorably oriented, if not strongly directed. I started school on my seventh birthday but I'm sure I was ready before, and within a year I was in the third grade reading. School was easy and I was the first in the community to go to high school. I liked all subjects in high school, particularly the sciences, and I did very well in all. In the senior year I took on the normal training course and read James' *Principles* and *Briefer Course*. It was in that course that I first got acquainted with psychology as a subject but I'm sure I was interested in it before that because as a child I made observations on perception which I later found out were psychological. I made observations on depth perception and I read about hypnotism as a child.

"Afternoons during high school I spent mainly in reading. I was never active in athletics. I don't think I had much aptitude for it. I think I was discouraged in early childhood. My brother being six years older was much stronger and he and his friends dominated the scene so that I couldn't compete. I was not interested in gadgets, although we had a good tool shop and I had plenty of opportunities to make things. I have never liked laboratory apparatus.

"I stayed at home for a year, and then taught grade school for two years before I went to the university. My teachers encouraged me to go and my friends were going. I don't think I had very well-defined objectives. The first objective was when I took a normal training course with the idea of teaching. When I started teaching it was with the idea of getting enough money to go on to college. My parents were favorable and helped me financially. I was planning then to be a psychologist, but I hadn't much idea of what I would do. I guess I knew that most of them were teaching.

"My course was interrupted by the war and more teaching and when I went back it was with the idea of being a chemist, but also I started psychology. I did sufficiently distinguished work that I was offered a job as assistant and this changed my vocational plans completely. That year the clinic was left without anyone in charge and I took over. I had just had one course in Binet but I dug in and read everything I could get hold of and I learned to give Pintner-Paterson and other tests. I learned a lot of psychology that way. I was offered an assistantship elsewhere for my Ph.D. It was a course in psychophysics that impressed me the most, the exactness and rigor. That is another reason for the mathematical twist. It appealed to me, it's neat and precise, and probably is the chief source of the direction I have taken since then."

G. "My grandparents on both sides were early pioneers in the West. Father ran a wholesale electrical jobbing house, and I was just about finishing high school when he was killed in an accident. Mother died the same year, just before Father. I had always worked in the business with him and I carried it on and then sold it after about six months.

"I had to repeat the first grade in school. I was pretty indifferent as a scholar. I've decided now I couldn't read in those days. But I think it's fair to say that I stood successively higher in the group from the first year when I flunked out to the year I got my Ph.D. I played outside after school until I was in high school and then I worked, helped out in deliveries and then played around with the kids in the neighborhood at nights. We used to roam the streets and break into the corner grocery store and steal things, like cars. I never got sent to the reform school, although I've been in the juvenile court for jerking trolleys off the cars. It was mostly good wholesome fun, although occasionally we were in trouble.

"I made a pretty good record in high school. I worked some on the school paper and I was on the debating team. Church was a very important part of our lives.

(After a period abroad he returned home and started to college, the expected thing in his community.) "I never took anything I was supposed to. Mainly I was taking many different courses. The tallest and broadest of the subjects were the ones that interested me. I had a good general education if poking into all sorts of subjects is a good general education. I started in law but soon talked myself out of that. About that time I looked into what this medical school deal was about and I was able to predict what room I'd be in at what time in the morning four years hence, and I couldn't face that much determination of my future. In my spare time as an undergraduate I began to hang around the psychology department a bit. I used to go up to the professor's house for seminars and that seemed kind of interesting. I read Watson. When I decided not to go to medical school I thought maybe I'd study psychology. Of course there was hardly anything I didn't want to be. It was a process of elimination. I talked to people and I remember one of them told me, 'Well, you go on and become an experimentalist and you can always have a good job.' I think I was helped a little towards this decision by taking the Strong which was coming out about then. I got myself scored on a few things and I got A in math and A in psychology and B's in law, medicine, banking and things like that, and way down in salesmanship, and that agreed with my introspection.

"I came East without bothering to get admitted in advance and shopped around among courses. I took a course in psychology and said I'd like to do an experiment. I was sent up to the attic where I set up an experiment. Then I had a great insight. I discovered that if you varied one thing, another one also varied. Oh! what an experience that was. I fiddled around and I had a bright idea how to compensate for this effect and I could plot some curves. Then I discovered about prelims and I had 100 days to prepare in, and then I was a psychologist."

H. "My father was a clerk in his father's general merchandise business. He had had some ambition to be a doctor but his father would not let him, with the result that he had ambitions for me to become a doctor. He was a rather good-tempered, relatively unambitious man, interested in politics and a definite extrovert. Mother was an extremely ambitious woman, and a very dissatisfied one. She undoubtedly had the feeling that she was better than the townspeople.

"I was able to read when I was four, and started school then. Then we moved and for two years my schooling was irregular but after that I was pushed ahead. Five nights a week at 8 o'clock I was put in a Morris chair with a board across the arms and I worked on my lessons and recited them to my mother until I was letter perfect. I was 12 when I went to high school and I graduated at 15. This was not an accredited high school and I had then to do a year of prep work. I was handicapped in sports but had not much interest in them. I spent all the time I could reading mother's considerable collection of books, mostly fiction. My only other interests were in boats and pets. Mother bought a parrot when I was two years old and we learned our alphabet together, and we were rivals in the family.

"I think I was an extremely passive, unambitious, and very obedient child. I was never punished by my father but was always in fear of him. I didn't discover what it was to have liberty until I got away from home. I was kept in long curls and short trousers to the point of making me ridiculous. This was because of mother's ambition to prove my precocity.

"I hated high school, the regularity. I have never been able to do anything seven days a week. I pretended to have headaches in order to miss school and I have been missing school ever since.

"I went to college in accordance with father's plans for me. I made very few friends and spent much time reading alone. I learned to smoke and I learned what it was not to have authority over me. I was disoriented for a long time until I found a library. For a while I had a chum and then in my senior year I was left again without close personal contacts. At one time I realized I was doing very little talking so I kept track of it and I found that in three days I had said twenty-four words. I took no part in class activities, I wasn't rushed for a fraternity. I had $28 a month to live on and I managed to make that go.

"I had no ideas about a vocation except some fantasies about civil engineering which did not meet with my parents' approval. It was decided I should finish my bachelor's work before I took any special training. I found I liked zoology and then concentrated on that, and I got an assistant's job in the laboratory. From that time on I lived in the laboratory. I developed quite a fixation on the teacher who became to me the ideal scientist. His method of instruction was to say to me 'You will find some sheep brains in the laboratory. Go get them and work up a course in the laboratory for the course in neurology.' There was a good deal of prestige value in this job. I became an authority among fellow students on biology, including the biology of sex in which I knew nothing directly. My only outside interest was in music and this has been a very important part of my life.

"I went to graduate school with a teaching fellowship in biology and I had no idea of what I was going to do. I was just living in the present, there was no future. I had a purely passive attitude but with this love of the things I was doing in zoology, and no idea of the possibility of a career in zoology. I have never looked ahead towards goals. It has always been, here are problems to be solved. I eventually majored in bacteriology but with psychology and psychiatry as minors. After my master's degree, I became an assistant to one of the psychologists to whom I became very attached, and I worked with him very closely. When I think about how things have gone it seems to me that the little success here and there seems to be a determining factor. That and personal contacts. I did some clinical work for a while but hated it and soon got back to animals. I've never been interested in people."

I. "My father was a physician and my own thought was of becoming a physician and taking up in my father's footsteps. This was a common practice. I was pretty timid about the problem of making a living. I had a feeling of being bright but impractical and this seemed a secure pattern. Mother always talked about wanting me to be a lawyer or judge, but I think it was her notion of what profession had the most prestige. We had a sort of *noblesse oblige* attitude toward democracy. We had a feeling of being somebody in our small town. I was invited to the right parties and all that. We were very good people. Father died in the war when I was fourteen. It wasn't a particularly sad experience in some ways for me, but it was sobering because I took a sense of responsibility, being the man in the house. But from then on the medical business seemed to be out of the question.

"In grade school and in high school things came easily for me. I skipped a good deal in the grades. I think this made me a little anti-intellectual in a sense. I wanted to show I didn't ever study or anything. I didn't read much, either, and I didn't do much with gadgets. I had some kind of psychological advice in high school, from two people who were sufficiently contradictory in their appraisal of me that it was a little reassuring in a way. The principal told me I was bright but timid and would probably never be comfortable with people so I should go into scientific work. So I decided to go into chemical engineering. It was kind of a romantic thing and it sounded a little hard-boiled. The assistant principal insisted I take up debating because I was sure to be a public figure. I was successful at this; in spite of timidity I had a good platform manner.

"I took my bachelor's degree in chemical engineering. I did well in my college work but I never really got very much identified with chemistry. I think the reason I stuck with it was because it was easy for me. I wasn't satisfied with it but in the midst of college life it was not so important. Other things, fraternity life and so on, were more important. I was very active on the campus, especially with campus publications and YMCA work.

"I didn't wrestle much with vocational goals until my senior year. There was pressure from the Y to go on with that kind of work because I was pretty successful. I was emotionally involved in these problems but hesitated because of the religious angle. My attitudes were scientific. The supernatural angle didn't appeal to me but the social did. The thought of an advanced degree never entered my mind, although my chemistry record brought me offers of financial aid. But I wasn't very successful in scouting around for work, which I took as personal rebuffs, and the places I did see where you could get jobs in chemistry were uninteresting and in places that smelled.

"So I took an offer to run the employment office at school and work as a Y secretary. It was a curiously reassuring experience that I could become a Y secretary without being threatened (in prestige). I was getting a little organized in a way. I received a grant to study at a divinity school and I had a thoroughly good time for a year. I was just exploring. I just absorbed it. I did a lot of reading in poetry and archeology, along with the standard course in religious education.

"In the meantime I had a job on the side in the student employment office, having some attraction to vocational guidance. I suspect that no one with very clear vocational plans would go into vocational guidance. One of my friends was writing a book and he needed some statistical work done on it, and I got Garrett out of the library and did correlations the next day. That was kind of interesting and I decided at that point to give psychology a try. I had only had a half course during my senior year. All I remember is being puzzled at the assistants and wondering why any one would ever choose that for a profession. I presently found myself in the psychology department there. I think by this time I was somewhat returning to the satisfactions of scientific work; I think the notion of apparatus appealed to me and of mathematical work. I liked my fling in the other world but I liked those correlations."

J. "My father was a lawyer with only a law school education and was the author of a book on compensation law. Mother had been a stenographer and secretary. They both grew up in the small town where we all lived until I was through high school. We all went to the same high school and father and mother and I all finished second in our classes. I simply went through the town school from first grade through high school. One of the teachers had a great influence on me. I had her for twelve years in different subjects. She read 'advanced' books like *Lord Jim* and told me about them. She was the only cultural influence in town.

"I wasn't much of a person for sports. I went through a phase of living out in the country, and later I played tennis a good deal, generally with older people. I learned to play the piano at seven and later played other instruments in band and orchestra. I read a good deal. Once when we were reading *As You Like It,* father casually referred to the Baconian theory and next day in class I said in a very smart-alecky way, 'Shakespeare didn't even write this.' The teacher said 'You don't know what you're talking about.' So that challenged me and I went down to the library and looked it up. I read a lot—all the Bacon I could find—and I must somehow have gotten the stuff under my skin because I've been a staunch Baconian ever since. I understood it later. At the time, of course, much of it was way ahead of me. I was always writing, too, short stories, poems, essays. As a senior I got a job on the local paper.

"Father wanted me to be a lawyer and pointed out the advantages of being able to come into his office. I never liked the idea but I didn't openly rebel, and I had no vocational guidance information whatever. It was taken for granted I would go to college. I always assumed that everyone tried to better his condition and I was much shocked in my later years in college to find that this wasn't always so. Probably my father being a self-made man contributed to that.

"I wanted to be a writer, but supposed I would go into law because I could do nothing else. I majored in English and minored in Romance languages and I went to college full of real respect for learning and supposed everyone else was doing the same. It was a lousy college. It was the most unplanned kind of education you can imagine. I took courses because my frat brothers recommended them. I took a course in embryology and I did a lot of extra work in it, some original research. I made slides and I had the feel of science. By all odds the great side of my college career was in learning something about the art of living."

As a tutor in a professor's family he learned much about art and music and fine living. This intensified his interest in writing and he was also profoundly affected by favorable criticism of his work by a noted poet, and came to feel that law was impossible. After college he stayed for a year at home trying to write but was unable to get started.

"During that year at home I read a good deal, McCurdy, Watson, Pavlov, Loeb, Broad, Ogden, and Russell. I was much impressed with behaviorism as an approach. I had had no psychology in college. I took biology because science was required, and now I decided I would study psychology. I wasn't equipped for any other science and I was always interested in literature. I decided that literature wasn't an adequate method of tackling the problem of behavior. I went on to graduate school. I was highly motivated that first year. I used to work from six in the morning until nine at night. Every moment was accounted for. That first year I didn't see a movie, I didn't have a date. I couldn't stand that pace now at all, but I did catch up. I was amazed to find that no one in the department was a behaviorist. I thought nothing else was possible. Since there were no behaviorists there, I went over to biology pretty much where I could see some behavior being studied. You become a psychologist to find out how to have your own way. You want to know how people behave as they do and partly because you want to change them."

K. "Father was a purchasing agent, but was also clerk of the board of education in this country town and clerk of the village trustees, and was greatly respected throughout the area. Mother had been a nurse. I believe that both of my parents were quite influential in the development of character traits. Both were proud and sensitive and extremely kindly and accommodating to everyone. Ours was not a demonstrative family; there was very little outward manifestation of affection, but a strong and unwavering sense of loyalty and attachment was understood by all.

"Throughout school I participated in athletics of all kinds. We all did a lot of hunting and fishing and were all the time in the woods, traipsing around somewhere. Father often took me with him on visits to different quarries. I enjoyed school very much, both the athletic things as well as school itself. Until about my class there hadn't been people going away to school. I think the main reason I went to college was the high school principal who was also our coach and was a great influence on us. He also taught the science courses and I think probably I got some of my interest in science from his being coach and teacher. My parents were willing but concerned over finances. They thought of college as a means to a business career, and so I majored in commerce as well as in psychology. I had all kinds of jobs. I ate one meal a day, but I was not alone in that.

"I was very fortunate. I took psychology my first semester. The professor was dynamic and inspiring. During the first year we could take the experimental lab and I had a good deal of interest right from the start. At the close of that year the professor's senior assistant was leaving and he offered me that job. I went on in psychology mainly because of the professor's influence. He became a kind of scientific father at that point. I was almost completely separated from my family as far as educational interests went and he would take me around wherever he went.

"I went on to graduate school with the vague idea of becoming a psychiatrist. Again I was fortunate in an inspiring teacher. I majored in clinical psychology and we attended staff meetings at the psychiatric hospital. I developed a tremendous interest in the physiological things. I think that's where my interest in psychiatry fell out, I felt I had a channel or avenue that was really my own. Those days were really very valuable ones for me, not so much because of the training there but because I caught the spirit and enthusiasm for research."

L. "My grandparents on both sides were small farmers, my father a skilled workman. When I went to school I had a lot of traumatic experiences. I had a hell of a time adjusting in kindergarten. But after the first year in grammar school I went on with not too much difficulty. I played with the kids after school. Mostly we played outdoors, although I occasionally did stay home and read. We played a good deal of baseball and there were some fights, as kids of course. I wasn't too aggressive but I got into fights. When I went to high school I didn't particularly get in with any group mostly because I worked a lot of the time. My family thought it developed character, it wasn't necessary. In my senior year I refused to work except Saturday and I graduated as first boy in the class.

"College had been projected, of course, from the time I was very young. I just accepted it. I guess I wanted it, I don't know. I signed up as a chemistry major. I guess it derived from father's interests. He had wanted me to try to get into a technical school but I didn't quite have the prerequisites. In college I took some part, not too successfully, in athletics, but what I enjoyed most was the Glee Club. And I was editor of the literary magazine.

"In my third year I began to wonder what I was getting out of all this business. I wished I had taken philosophy because I was concerned with problems of the soul and this bothered me, but there were no philosophy courses and so I took psychology. Although we were taught Titchener pretty much and although even then I didn't believe it (I thought all of this could be described in physiological processes) it seemed to make sense and I liked it. It gave an experimental approach to problems I was interested in in philosophy and a better approach than sociology did. I decided to try to get all the psychology I could and though the experimental course was very tough I took it. When I was a senior I worked with a Ph.D. student on a laboratory problem. I thought this was really the stuff in the sense that this is serious as the experiments in experimental psychology were not, so I got quite a feeling of responsibility out of it.

"I went on to graduate work. The family didn't object. The professor left students to themselves not giving very much aid. I half liked it and it half irritated me. It developed a sort of anxiety in me. I would see some of the other people going ahead and getting something and here I was reading at random still. During my first year it wasn't much fun. I guess I worked a good deal alone. I spent the year studying mathematics and science generally. There was no emphasis on psychology. All of the students were individualists; there was little homogeneity of interests at all in the whole group."

M. "My father was an engineer and had a polytechnical training in Germany. He was killed in a hunting accident when I was a child. Mother took over his business, and from then on we lived in boarding houses, and I was left pretty much to myself.

"I was always a very poor student in all subjects except math, and I could do special problems, but I didn't do too well. I always just managed to get through. I spent a lot of time in my father's shop. I was a little pet around the place, and they let me do everything. I've always enjoyed working with my hands. I still would rather do that than eat.

"I spent a lot of time in athletics; I was in all the track events in high school and football and baseball. Father's plan for me had been a technical education in Germany so after high school I went there. Suddenly the world opened up, the world of literature and all the rest. There were bookshops on every corner. I read all of the Russians and then I read all the German literature and then I went back to English literature. I became very much interested in the theater. The first year I was sort of outside the picture and I was terribly lonesome and that gave me sort of a chance to look back and see what it was all about. After the life I had in high school where I had been in everything, president of the class, captain of the teams, and then going to this life where you didn't have to attend any courses. The second year I was elected to a sportsverein. I was there three years, until the war broke out. That brought about a general re-examination of what you wanted to do and I suddenly found that engineering did not interest me enough to consider it a life career. I didn't know really what I wanted to do. I was interested in something that had to do with people and behavior in more general ways. As closely as I can remember it was a choice between economics and a diplomatic service career.

"I came back and went to college. One course in international diplomacy cured me of an interest in that, and I soon gave up economics courses, there was nothing I could hold to, it was all verbal. Then I got interested in Freud and early analytical work and I read everything in analysis. I had one course in animal behavior; this was my first contact in anything like that and I got an A. I knew then I wanted something in the form of psychology. I had read Watson, I was stimulated from the physiological angle. I got my B.A. and when I got out of the service I came directly here from the camp and started in. This was the best place I could have picked out. The professor said just go ahead and work, there will be no courses, just go ahead. So I went right on, on almost the same system I had in Germany. I made arrangements to dissect a stiff in anatomy, I just did it on my own and I did the same thing in physiology. I ran all the experiments on my own and then that was the extent of my formal education. Meanwhile I did a lot of psychological experiments. Actually so far as formal education goes I had very little. I just managed probably to be lucky and not to have been forced or unlucky. I've been able to go ahead in developing my own interests. I was a poor student, I could never have made the grade in any formal course.

"I had no idea of a career at this point, it was just something that was overwhelmingly interesting to me. My professor had a very important influence in my whole development, he was far and away the most widely-read man, the greatest scholar I had ever come into contact with. He had a great fund of information and high ideals and tolerance of all kinds of work. As the work developed, as I look back I have been stubborn and unpleasant because many times he tried to push me in another direction. He had little idea of what could be worked out experimentally. I got my degree in two years, and after I graduated I had a staff position. It was very exciting to have contact with the patients. It opened up all kinds of possibilities."

N. "My family background is both farm and fairly good upper-middle-class group. They had farm, mercantile and professional affiliations, medical and legal. My father is a very able, hard-working man who put himself through college after he married, went on to graduate work and ultimately became a professor. We moved around quite a lot. I never was particularly interested in school. I was one of the good boys and I got along nicely in school because I was good. I was one of the rather non-social shy children whom teachers all like. My scholastic interests were minimum, except English and drama.

"I didn't get along too well with other children, but always had one or two close friends. I had one difficult year when we moved and I was a stranger and very tall and thin and physically ineffective. I found out what it meant to be a minority group member. And then we moved back and sex had arrived, which was very happy. My high school years were very highly heterosexually oriented with lots of dating and dancing and great interest in reading and writing. I did a lot of acting and journalism; I played the piano and learned to play jazz. I was thoroughly and completely an ingroup member then.

"It never occurred to me that there was anything anyone did except go to college. I don't think vocational plans ever entered my head. By my junior year I decided to become a short story writer and became quite a Bohemian, interested in the esoteric and an expert on metropolitan speakeasies and on local wines. At the end of that year I got engaged. She'd had to go to summer school so I did too and we took Psychology 1. We sat in the back row and held hands and went out and studied together under the trees. This undoubtedly had a profound effect on my interest in psychology. On the other hand, we also took sociology together and it didn't have any effect on that. I looked on psychology as a refuge from the vagueness and what even then struck me as the amateurish guesswork of English criticism. I was quite dissatisfied with the balderdash about motivation and character and so on that you got in that. By that time I had begun to suspect that I might not be the kind of short story writer that eats. I liked learning theory and I loved the objective questions. Just this incredible number of isolated little facts. It provided a certain solidity, it was so nice to get things down to a really precise point. And I found some very smart, sharp, good people in psychology who were very much interested in me and who spent time with me and gave me interesting things to do, like measuring things and finding relationships. If any one course were to be given credit for my final choice, it would be the course in experimental.

"I went on for graduate work with the notion that I would go only for a year and it would be nice to get far away from home and fun to see something of the rest of the world. I suppose this was probably about 80% of a decision toward a profession but I can remember I had a feeling of not having made a decision. I couldn't seem to think of any alternative. I didn't like it, it was ghastly, but after two months you couldn't have gotten me out of it. I guess it was like taking religious orders. Part of it was the enthusiasm of some of the people. I started working with one of the professors on an experiment; I did a lot of research and then there was just no further question."

SUMMARY

This group came from lower to upper-middle-class backgrounds and the economic level varied from quite poor to well-to-do. Many of them had feelings of apartness relative to themselves or their families but it is rare for these to be colored with inferiority feelings. More than half of them had some definite sense of personal or family superiority, and family concern with social status, in one way or another (as striving, as recollection of striving in the parental generation, or as consciousness of belonging to the "best people").

The earliest vocational decision for any of this group was the sophomore year in college, and over half did not decide until after they had graduated from college, several not until they were part way through graduate school. Their earlier interests were oftenest English literature, although some had social interests and some began in chemistry or engineering. The final deciding factor was often the experience of doing research.

There is no consistent pattern of reading interests during childhood and adolescence. The amount ranges from none to "reading was my life," and there is no concentration of frequencies. In college or high school, six of them did some writing or editing or both.

School work was always easy for most of them. Only one said he never did well in school, but several others did not make exceptional records. Two were extremely interested in school athletics.

Four suffered the loss of their fathers (at 8, 12, 14, and 17 years), one also his mother, but there were no homes permanently broken by divorce or separation. One subject's parents were divorced but remarried shortly afterwards. The father of another died while the son was in college.

Inquiries about health during childhood and later uncovered a number of problems, of varying sorts. There is no record for one. Five stated their health had always been good but one of these developed migraines when about 25 and another developed a number of allergies in later life. Two have had hearing difficulties from an early age, and two have had some eye problems which have been corrected. None of these problems seems to have been an important early factor.

There were 4 who had quite serious problems which were constitutional. Two of these were abnormally small and 2 were abnormally tall. These were their comments: "I was always pretty small and slim but my coordination was not too bad. I made basketball (in college) but dropped out, I was too small for one thing." "I was always a very tiny boy. When I entered high school I was 4 feet 10 inches and weighed 68 pounds. I was small and so the group I went around with was younger. In college I had gotten to normal height but I always had this picture of being minute." "When I was 15 years old I was 5 feet, 11 inches and I only weighed 95 pounds. I was incapable of doing most of the things required in gym, so I used to get under the mats or hide behind them and stay there as much as possible." "I was very tall and very thin and my adolescent growth spurt had come at the beginning of 12 years so I was very ineffective and I had sick headaches. I suppose that they were probably excuses to get home from school. My health has always been doubtful. I am somewhat hypochondriacal." One of this group was always an isolate but the size factor was a relatively minor one in this picture.

There were other problems. One subject had diphtheria at 4, followed by a temporary partial paralysis, but was otherwise never seriously ill. Another was sickly as an infant, had many serious illnesses as a child but had not had any since. A third suffered from one long series of stomach aches and nosebleeds, probably because of poor nutrition. Another said, "I was fairly sickly when I was a kid. It was a very rare year that I didn't miss a block of 2 or 3 weeks or more of school because of some illness. And I must have had ulcers by the time I was 15."

It would appear that health and constitution have not played any clear role in this group generally. In only one instance does it seem likely that they contributed significantly to difficulties in social integration.

The patterns of psychosexual development are discussed separately, in Chapter V.

THE ANTHROPOLOGISTS

O. "My family background is nationally very mixed, but it is mostly farming, although father was a small businessman. We were definitely in the upper-middle class until we moved when I was 12, and our status changed to lower-middle class, which was very painful for Mother. I had a very unpleasant time in grade school. Mother insisted on status differentiation in terms of clothes and so on, but after we moved this situation changed, and I got into athletics, too. At studies, of course, I was a very bright youngster, and that was another thing that put me at a disadvantage at an early age as I was considered teacher's pet.

"Pretty much from the time I was 7 until I was 15 I was quite determined to be a writer. When I went to high school some one had tipped me off that if you want to be a writer and eat you ought to take courses in journalism. About half-way through high school I don't know how it came about, I decided I wanted to go into archeology. I knew very little about it except the King Tut tomb stuff. I had visited the Chicago Museum, and I think I heard my father talking about it and I had read something about it. I suppose the aspect that appealed to me was that it combined both the rugged outdoor life, this side that I was always striving to be proficient in, with some intellectual content, and I couldn't just be a ditch-digger but this was ditch-digging with an intellectual content and it was glamorous.

"College was always taken for granted. The school advisor didn't know that anthropology was also archeology, so she could only find a small state university that seemed to give it as a major so I went there. Many of the students spent quite a bit of time mending pots or working around the archeological laboratory, but I never did. I would never go on field trips during vacation because I wanted to spend my time on the beach. Looking back over it, it looks rather bad, as though I didn't have any honest interest. I must have been pretty much of a horrible social snob, too. The other students in archeology seemed a little second class socially to me, and there were times when I thought maybe I'd do better to switch back to my earlier ideas of being a writer and I took as much in English as I did in archaeology. During the year I worked for an M.A. I did quite a bit of reading and I became really serious. The chairman of the department encouraged me to go on, somewhat to my surprise. I got a summer fellowship which opened up a wonderful door for me. If it hadn't been for that I would have been out of luck. I learned a lot and I learned how little I knew. I worked for two years with a man I met on that trip and for the first time I got to know something. Then I finally got a fellowship and went on for my Ph.D."

P. "There were no professional people on either side of my family and my father was a superintendent for a building company. We were not well off, but mother had been a teacher, and she kept up pretty close personal contact with all my teachers. I did quite well in elementary school. I think I was a pretty conforming fellow, all the motivations were stacked up and mother was interested in grades. I went to a manual training high school because they had a swell mandolin club. I wasn't athletic but I worked on the school paper, and I had a wonderful time and all that kind of thing. At the end of my high school career the college question came up. My parents wanted me to go and I didn't care. I didn't know what else to do. I went to the business school because this didn't require passing exams. I took a major in economics and spent a lot of time with the mandolin and glee club.

"Through contact with social science, the problem of social reform began to agitate my mind a bit. I took no anthropology, I had never heard of it until I met an anthropologist who was a member of my fraternity in my senior year. By the end of college I wasn't interested in doing graduate work in economics and I didn't want to go into business. I had done some volunteer social work and I was offered a case worker's position. Everyone thought it was a funny thing for me to do at that time. After a couple of years of this I decided to do graduate work in sociology and did so while I kept a full-time job. I took some work in anthropology and I discovered, and this stirred me up intellectually, that as in contrast to sociology, anthropologists were very skeptical about social evolution. That intrigued me, so I began to read and I got more and more interested, and dropped sociology altogether. The department was small. The professor was the kind of fellow that loved to have people hanging around and talking to him, so I got a lot of stuff rubbed off outside of the regular classwork, and he wasn't concerned about attendance.

"I was still in social work when the psychological revolution began. That's when I began to read in psychoanalysis. There was no psychological aspect to anthropology at that time and I had taken only one formal psychology course in my life. I gave up social work when I was offered a fellowship. By then I had determined to make anthropology my career. At that time I had done no field work, but was soon able to make a few short trips. When I began to do field work there was practically no psychological slant at all. It's hard to make a precise statement as to how I first brought an old interest in personality psychology into relation with the study of culture. I knew A. A. Goldenweiser, Ruth Benedict, and Margaret Mead personally and read all of their publications and I had listened to Goldenweiser lecture on psychoanalysis. Then there were the early programmatic articles of C. G. Seligman and Edward Sapir."

Q. "My father was a business man who did a good deal of foreign business and was away a great deal but sometimes he took me with him. My mother's family came over in the Mayflower. They were so bogged down in the haze of myth that I can't make out much about them. My grandfather took care of me. He told me all sorts of stories, so I was brought up on a diet of bloodshed and adventure when I was a child. I didn't think much of that until recently but I think it's probably important.

"I went to regular school but I was always getting expelled and sent home. Then I used to go out in the woods all the time. I got into all sorts of trouble in high school. We made every possible attempt to outwit the authorities. I led a stink bomb attack on the faculty and then father sent me to a private school. I graduated *cum laude* and took first prize in Greek. I was doing a lot of reading of Egyptology and started learning hieroglyphics but I couldn't see that what I was studying had any relation to what I wanted to do, so I flunked plane geometry three times. I didn't like anything except geography and Greek. I couldn't do math and I hated history. Languages were a cinch. I wanted to be an archeologist and an explorer. I read adventure magazines and all those things.

"I figured it was necessary to go to college, and Father picked the place. I wasn't particularly happy, I didn't like it particularly and the professors scared the life out of me and I got jaundice. I started out to major in the classics but I was disappointed because I discovered the reason I liked the classics at prep school was the teacher. But from then on I was taking mostly anthropology and I was getting around with the graduate students mostly so I was much happier. I spent several summers abroad. We used to see how many borders we could sneak across, and I got in some field work in anthropology."

After a half year of graduate work he went into the field on a fellowship where he had an assortment of physical mishaps, and illnesses, which put him to bed for some time. However, he managed to do some studying and got through his prelims. Later he went back to the field to work on his thesis, and had his usual exciting time. Again there were illnesses but he eventually got back and took his degree.

R. "There is some intellectual tradition in my mother's family but none in my father's, although there were some professional men. Mother died at my birth, and when I was about 5, I went to live with an uncle who became my foster father. He had inherited sufficient money that he did not have to work, which is not the custom in a middle-western town. There were two distinct social groups in the town, and father was accepted into the more snooty one, but mother was not.

"In grade school I was very successful as far as grades were concerned, and I had one year in the local high school. I did a lot of reading, and I had an active social life and lots of athletic activities. When I went to a private school for my last three years I was young for the group and too undersized, and also rather naive. I wasn't any athletic shakes and this was terribly important. That first year was very bad, but I returned out of sheer stubbornness, and my last year was very pleasant. I was too young for college so I was sent to another prep school. There, in many ways I found myself. I found a kind of life in my own age grade which more or less inchoately I had been groping for. A kind of realization of some of my intellectual and literary efforts, and in contrast to before, when I entered as a young squirt, here I entered at 16 with military school experience and a little athletic renown and I got away with being something of a sophisticate. The more I liked it there the more I immediately started to look down on the first school in intellectual things and I'm afraid in class terms I was something of a snob.

"College started off very well but I soon got into rather involved difficulties and was sent home. After some time I went West to visit a relative. This is a nice point of the role of accident in human life. If my mother's cousin had had a ranch in a different place or had been without intellectual interests, the whole future course of my life would have been very different. This guy had intellectual interests and a good library and lived in Indian country. Without realizing it, this was inevitably the beginning of my anthropological interests.

"Eventually I went to another college. I got along well. I was elected president of my class. I fell in love and I finally decided to finish there. I was thinking of the law but only of an undergraduate degree in it. I got away with murder there, I had a good memory for words and verbal facility and while some teachers were sincerely motivated and deeply impressed me and what I did for them was honest, a lot of my instructors I just used to twist around my finger. I would throw esoteric references at them. I did crazy stunts that made me notorious. I was president of the male student body and chairman of the newspaper. I was out of athletics but I led a pretty vigorous social life. I thought about literature as a career but some of my teachers had convinced me that while I was not completely hopeless, I was not likely to set the world on fire.

"During the summers I took pack trips among the Indians with friends. I didn't know there was such a thing as anthropology. Then I got a scholarship abroad and that was quite an experience for me. I had had an experience of four years of being a big shot . . . for the first six weeks abroad no one really spoke to me. I couldn't get away with anything in my tutorials and I was very much humbled and I realized I had been a fraud intellectually. I decided to stop that, so I worked far harder than I ever worked in my life. I just worked like hell. After the first year I had to some extent caught up the lost ground so I could hold my own.

"I came back and tried a few weeks of law school. I thought and discussed things and I did quite a bit of reading and finally I couldn't tell you exactly when or how, I don't remember there being any sudden revelation about it but somehow the word anthropology had come to me and I had done a little reading, precious little, in the field and I made up my mind that I was going to become an anthropologist. I did this for several reasons. It would give me a chance to be out-of-doors which I loved, and it would link my life respectably in a socially approved way as opposed to that of a dilettante with the Southwest and then I remember saying, 'If I go into anthropology I can study any damn thing I want to and it will be part of my work.' I meant anything related to human activity and it will be part of my work."

S. "Both of my parents were born in the West, of pioneer families. My father was a lawyer with a good practice. The men in the family are generally professional and mostly lawyers. Father was raised next to an Indian village and was attorney for a number of the Indian tribes, and of course I got interested in the Indian background.

"I went to a public grade school, then to a private high school until my last year when it failed and I returned to public school. After school I went home and played around the house with neighboring children. In high school I went into the woods and studied birds. I did an awful lot of reading. The house was full of books, my father loved books and I read everything in the public library, particularly nature stuff and military stuff, nature books and travel books. I read some dime-novels, of course, but I don't think I ever went through a period of trash.

"I don't think anyone thought I would go to college because I was very poor in math and languages in high school and excellent in history and English and things I liked, and I think the family thought it would be a waste of time. But I wanted to be a naturalist. I didn't know much about how to become one except that I knew I would have to go to college. I was interested in ornithology particularly. Father was pleased but he said he had never heard of anyone making any money as a naturalist but that it was all right and if I became good at it he would help me if he had any money.

"Then in my senior year in high school I enlisted in the Navy. That experience is important. I think I would have made a good officer, but I did not make a good enlisted man because I was not mechanical. The main thing I was travelling, but we didn't see anything of the world except the ocean. There was an old-fashioned but good library aboard this ship, including Darwin's work, and I read the *Origin of Species* and the *Descent of Man:* that made a great impression on me. It gave me some scientific background for what had been just a collector's instinct and innate love for nature. It was in the Navy I began to see a difference in a very marked way between the officers and the crew, the lines are very sharply drawn. I began to see that it wasn't just chemistry and math but that there were other things that you could get and learn.

"I registered as a pre-legal student and later changed my major to history. I didn't know anything about taking biology. You can't take a course in Darwinology and I didn't know what these other things were. I collected birds for the college museum, and then in my junior year had a field trip with a museum party. That was fascinating. I was doing professionally what I had done by myself and became aware that people did it as a business and at least were able to eat some of the time. Then I had to find work and the professor of anthropology gave me a job which involved real research. I became more and more interested and then I found to my surprise that he had put me down as joint author of the reports. Once I really got into anthropology there was nothing more for me. I think that if any motif runs through, it's that interest in history because I went from law to history, it's history that I was interested in in high school and the major aspect of evolution. I had the interest of a naturalist, too, but dissecting left me cold, I was never tempted to go on in zoology. My three graduate years were without doubt the most exciting of my life. It was a period of life opening up as more than work, and sports and books. We were all poor and working like hell."

T. "My family on both sides were farmers with a fair education and of moderate social status, and I was born on a farm that had been in the family many generations. My father was a man of great ability, unusual intelligence, and violent temper who made and lost at least 3 fortunes. In the days when I was growing up the family was comfortably fixed but not wealthy. I was never close to anybody as a child. I went to a small religious school and there I was intensely unpopular." (After some years of being bullied he developed enough strength to fight back and in a deliberate campaign thrashed one boy a day.)

"From the time I was 10 I worked in my father's business holidays and summer vacations, although later I took summer farm jobs. In high school I was an omnivorous reader. The family subscribed to various current magazines, like *Harpers,* and there was a library in town and I got books from it. I was very eager to learn and there wasn't much chance to learn. I found elementary chemistry and physics mildly interesting, I think the reason was that it was the livest research subject. I learned about research from the old *Scientific American.* I knew about it very early and I knew I was interested in it, but the education in the schools was very poor.

"I went to college planning to be a chemical engineer. In school I'd never had to do any studying, the general courses were geared to a much lower IQ so that it never made me work and I had no idea how much mathematics would be required. I've never been able to do math. I did no work my first year in college; out of 34 credits I flunked 11 and was conditioned in 17. My father said he would give me a job but that was the last I could expect from him, and I had completely fulfilled his expectations. Of course this made me mad and I made up my mind that I would put myself through college, so I went back and worked my way through. Then I decided that if I was paying for my education it might as well be something that I was interested in and I went into biology, which was the only other science they gave. There was a period in here when I had difficulty making up my mind whether to go into writing or biology. It was not entirely youthful enthusiasm because as an undergraduate I wrote well and I had several things published. When I took up biology I was thinking of the great open spaces, of collecting and exploring. I had one course in so-called anthropology and in my sophomore year went out on a field party, and a year later on an expedition. By my junior year I had made up my mind, and after graduation I went right on."

U. "My father was a well-to-do businessman and we lived in suburbs and had private music lessons and other advantages of that kind. I liked school and did very well. I began an avid reading career as soon as I was old enough to take out a library card, and my other major interests were drawing and music. I went to two poor high schools and then to a classical high school with extremely high standards. It was really a delightful school with definite emphasis on scholastic achievement and to hell with athletics. You were expected to do three hours a night homework and sometimes more, and the pressure and competition were hard at first, coming from a regular school. There was lots of intellectual snobbery. Of course there was an excess of bookish children so I was happier there than I would have been elsewhere.

"There was no question about my going to college, it was always assumed I would go. I tried to get into Annapolis but was unable to pass the physical examination. It was a terrible blow, I was very depressed and upset. I know now I would not have liked it too well. The idea of the freedom the Navy seemed to offer in getting around the world was probably one of the factors in my great interest in far places and in field work. I had no vocational plans when I entered college, although I had thought vaguely of being an author or a doctor. I started concentrating on classics but toward the end of my freshman year I felt I had to make a change. I was very snotty about English because I had already read so much that was required and I thought no professor could teach me anything at second hand. I had read Darwin and I had read a little about anthropology so I took a general course in anthropology and loved it. I really enjoyed every bit of it. My interest was almost immediately in the direction of the biological side. Of course physical anthropology is the most precise part. I like things that have a clarity and a precision and I like the concrete aspects. I like having the materials under your hands. I took a course then in physical anthropology, still not thinking about doing it professionally but I was very deeply interested. I had had a taste of business experience summers and decided against that as a career. In my senior year there was another thing. In the usual arrangement you take courses and the professor comes in to lecture and you never see him outside and this is one thing, but working in a museum where you are part of the apparatus of the department and you get a taste of what they were doing privately, where you see the work, this set-up might very well have been sufficiently entrancing without my quite realizing it. It's the first time I have thought of it. I vaguely knew that people could do research but that I could do it hadn't occurred to me until then and it was a very exciting prospect. By my senior year I was well set although vague as to what and where. But I got a fellowship for field work and after that there was no question but that I was headed for a career. Of course there was the whole mental turmoil and excitement and the drive that comes with a great absorption, that's understood. I'm

V. "I was conscious from the first of belonging to two pasts because my father's family on the whole represented an old American family that had lived in the same place for four generations while my mother's family were all Europeans. The family was one which was more self-engrossed than most at the time in America. We developed forms of living which were different from those around us. I had a number of serious illnesses as a child and father was very protective. I studied with tutors until I was in the 8th grade. I made serious natural history notes and read a great deal. I had lots of opportunity for individual experience in following things out just because I liked them. As I grew older the difference [between him and others] became one of superiority, but along with this went the feeling of superior responsibility. I could do things other people could not do and therefore I had to do them. I knew early that you could have a career as a biologist and that was my first plan.

"I had some difficulty adjusting when I first went to school but in high school there was no awkwardness. My interests then were chiefly literary; I had no particular interest in laboratory science but I continued to do natural history. By this time I was writing poetry and so were a good many of my friends. We were a somewhat precocious group of literati.

"My college career was interrupted by the war. I enlisted in an ambulance corps and I took a very bad beating. The group was a nice group to be with, but they were a bunch of hoodlums, undisciplined, and the occasion was such that they didn't get any discipline. I first came into contact with the way young men really lived. But I saw hard service also. Then the unit was disbanded and I just came home. I came back very much confused and disorganized and this lasted for some time. I didn't know what I wanted to do.

"I finally started law school and was admitted to the bar. I got to a state of great restlessness. I didn't like my work. My wife is an enterprising person and willing to take chances and with her urging I took a long vacation. We went to a primitive area, and through an old acquaintance from there we met an anthropologist working nearby and we became intensely interested and did some field work. Meantime I was greatly encouraged by a professor of sociology and through his influence I became very excited about concepts of the science of society. So I became an anthropologist and went back to school and did more field work. My personal life was opening out very richly and it was wonderful."

SUMMARY

The average economic level of the anthropologists is clearly higher than that of the psychologists, and concern with the social status of the family or a firm conviction of the social superiority of the family is evident in all but one instance. This did not always result in the development of a definitely socially snobbish attitude in the subjects, but there is good evidence that most of them did consider themselves superior in one way or another. All but two of them went to private schools, either elementary or secondary, and this would certainly tend to foster these attitudes.

Two of this group decided on their vocation in high school, three in the later years of college, and three after graduating from college. Their earlier interests were somewhat varied. Four mention outdoor and athletic interests as having been of importance and a factor in the choice of a profession involving field work. Two with natural history interests, one with an interest in science generally, and three with interest in the classics could continue to find these interests in their profession, as could those whose earliest professional interests were sociology and law. Six of this group at one time or another had a special interest in literature and writing.

Reading interests are varied. Most of them did excellently in school, or at least did so in those subjects in which they were interested, and characteristically put little effort into the others. They have, however, a general dislike for mathematics.

Health during childhood and adolescence was apparently good for only three of them. Another had good health until an attack of rheumatic fever during secondary school, with some sequelae, which have not interfered in his field work. There are five who apparently had constitutional difficulties. Three were underweight or undersized, and in addition one of these had a number of allergies and the other had a number of serious illnesses, sufficient to have affected his early schooling. One was oversized ("I don't know whether it was pituitary or overeating, because eating was about the only satisfaction I had"). Another was always the tallest in his age group which sometime gave rise to awkward situations. Another said, "My mother or at least I, had the idea that I was always a sickly child and I was always having to go to bed but there was nothing really wrong with me."

As in the case of the psychologists, the importance of the discovery of the possibility of doing research as a factor in choice of vocation is clear.

Comparison with Other Scientists

Although there is not much difference in the general socioeconomic background of the different groups of scientists (except for the subgroup of theoretical physicists, 84% of whose fathers were professional men, as contrasted with about 50% in each of the other groups), there does seem to be a difference in their social attitudes. Among the biologists and physicists I encountered no direct expression of feelings of personal superiority, and there were very few by inference. One of the physicists did say that the family considered themselves extra privileged in spite of their extreme poverty; there are a few others who probably had some vague feelings of family superiority on one basis or another, and there are some who were conscious of their intellectual superiority, but they seem not to have translated this into social terms. It is, of course, not certain whether this is because these groups don't think in such terms, and hence it would not occur to them to mention it, or because they actually do not have such attitudes. I think it is primarily the latter, although the former may play some part in it—it is an aspect of their rather general indifference to or avoidance of personal interaction. But among the social scientists, in at least half of the psychologists and in most of the anthropologists, a feeling of social superiority has definitely played a role in their development. In some instances this feeling is a product of the family's or particularly the mother's strivings (or a paternal grandmother's). Further data on this point will appear in the next section.

In the matter of early interests (the term refers to spontaneous activities) this group differs markedly from the physical scientists, almost all of whom displayed early interest in mathematics, chemistry, physics, or gadgeteering, and very few of whom were ever interested in literature or the humanities. Two of the psychologists and one anthropologist began in chemistry but quickly shifted. Literature and the classics, and less frequently social welfare interests, were common among both anthropologists and psychologists, as were some natural history interests, particularly among the anthropologists.[3] The biologists included men whose early interests had been in natural history, in literature, and in chemistry or physics, although the latter interest seems to have been aroused largely because these were the only sciences available in high school. In the histories of the social scientists and of the biologists the importance of the discovery of the possibility of doing research is highlighted, and this was often the factor that gave the final determination to their choice of vocation, or that fixed them in it once it was chosen. This particular aspect did not appear among the physical scientists, but this may well be because the difference between gadgeteering and experimental work is really a matter of degree and emphasis; the possibility of doing things yourself is obvious, whereas in the other fields it is not. It would seem that this may be an indictment of the pedagogical techniques in general use.

For the total group of scientists the median point of decision on a vocation is in the later undergraduate years, but it can be as early as high school and as late as postgraduate years. It is later for the social scientists and particularly for the psychologists. Psychology is encountered late in school, and lacks the popularization given King Tut and other archeological stories, and the adventure aura. But very few in the total group did any long range vocational planning.

Among the biologists, 5 lost father or mother before the age of 10, and the parents of two others were divorced (when the subjects were 9 and 16). Among the physicists, 5 lost a parent by death (at ages 5, 6, 9, 15, and 17) and the parents of one were divorced. There was only one divorce among the parents of the social scientists (and they remarried), but the mother of one anthropologist died at his birth and 4 psychologists lost their fathers by death (at 8, 12, 14, and 17) and one also his mother at 17. In the case of the biologists and physicists where the losses occurred very early, it seemed possibly to be a factor in the acceptance of isolation by the subjects, but among the psychologists and at least one of the physicists whose losses were later, the effect seems to have been more one of increasing the problems of adolescent reaction to authority, and this effect seems to have been greater in the case of the psychologists who have been more concerned with personal relations from the start.

A special factor, occurring generally only in the theoretical physicists, was the apparent effect of severe childhood illnesses which contributed to personal isolation. In all of the groups there are a number who had developmental problems related to constitution—abnormalities of size or general weakness. Unfortunately I have been unable to find comparative figures for the general population.

---

[3] Baas' study of interest patterns of psychologists on the Kuder showed that all psychologists groups had high scores on the literary as well as the scientific scale (2).

V. PSYCHOSOCIAL DEVELOPMENT

In this chapter an attempt will be made to summarize for each of these subjects something about the family climate and discipline, and their relations to their parents, as well as their general social development. There is a good deal of information for most of the subjects. Each paragraph refers to a different subject; the paragraphs have been arranged in groups to illustrate special situations.

One subject lived away from home during much of his schooling and was off on his own when he was 18. His parents seem to have been thoughtful and helpful but there is no feeling of closeness. He grew up very much to himself and was early discouraged in athletics because he could not compete with his older brother who "dominated the scene."

(A number of the families were self-isolated. These include the next 5. In the fourth and fifth the isolation resulted from the mothers' attitudes of superiority.)

In one family relations with the other relatives, as well as with the immediate family, were very close. "I never saw my father and mother quarrel. It was a completely false family picture of idyllic bliss. I think the family were completely protective and it's always puzzled me as to why. Mother was very close. I always thought father was tremendously fond of his family but completely unaware of them. He definitely dominated the family. He would read at the table. They had no social life of their own. The family was essentially socially self-ostracized."

Another family pattern was "kind of incredible. We were a family that kept completely to itself. They were very adept at thinking up things for us to do. They seemed so good that it was very hard for the children to rebel. The notion of not coming directly home from school would never have occurred to any of us."

The family of another were also self-isolating, in part because of living in a country place much of the year, and in part because of the personalities of the parents. "Mother was intensely sensitive and extremely fastidious. My father was intensely protective of his family. His idea was to save them all trouble, to keep them all from harm, to keep them secure from the world and to provide them with material comforts. The only conflicts I had with my father were over this. As I got into my teens my form of reaction was to become uncomfortable at the overprotection. I can only remember a few instances where there was any open conflict." But he enlisted without telling his family until afterwards.

One reports, "I was never punished by my father but was always in fear of him. A threat of punishment was always held over me. Mother used a fire shovel on me. I had a considerable mother attachment and great antagonism to Father. He never gave me any explanation of anything. He was completely authoritarian. I took it as a matter of course until I went away from home."

Another says "Mother carried embittered feelings about how people had looked down upon her because of the poverty of her childhood. She looked on her marriage as a vindication. She was a very strong-minded woman and was dominant in the family. Father is quite a passive person. I was always in conflict with her but there was a great attachment."

(Those whose fathers died when they were quite young are grouped below.)

Discipline for one was "liberal but with fairly rigid standards, but coercion was subtle. We were all good children, we never had a hand laid on us. We were indulged, we never did any chores. Mother kept us dependent on her in some ways and she could get us to do what she wanted us to. We developed a pretty strong sense of moral oughtness."

After the death of the father another family moved about a good deal because the mother had never established any roots. The early period had been very different, although the father was away a good deal. "You have the impression of this strong dominant father coming home and everything getting organized and there would be dinners and so on and then he would leave and things would quiet down." He went through a difficult adolescence with generalized rebelliousness, which got him into trouble at school. "It probably should have been more directed against Mother. It would have been more relieving but it could not be because she is a sort of unstructured person. The death of my father was crucial because it came when I was working out authority relations and I had no one to take over as a surrogate."

Before the death of his parents another says he was much closer to his mother, although a long illness kept her very restricted. "Mother was very popular and sweet-spirited. Father was quite taciturn and strict in a sense and kind of a compulsory, demanding person. He was strict in the sense that I would get hell for not taking care of my bicycle; I never felt I was picked upon. I respected him. His weaknesses are my weaknesses. I don't know what would have happened if he had not died at that time. I remember taking his reprimands more seriously when

I was younger. When I got to high school, we regarded each other as individuals."

One subject spent a good deal of time at his father's place of business until his death. "I think I took his death fairly naturally, although I was upset of course. And it meant Mother had to step right into the business and I was alone a lot and had to shift for myself. We hadn't been close but I had spent so much time in his business and we did a good many things together. Mother and father had violent disagreements and I was bothered that my father was a rigid person and I must have been beginning to rebel against that set way of doing things. Of course then mother and I were very close for many years. But she let me go off to school, she never restricted me, she never seemed to be jealous of the things I did." Since he did a very good job of shifting for himself, and was very popular in his age group, the net result seems to have been that he avoided a serious conflict in adolescence.

(In the next few, there were no overt problems. Note that two make the point that their being away from home eased the situation.)

One says that his father had a very mild temperament, and he does not recall ever seeing him angry or being particularly disturbed over him or any of his sibs. "We had a great deal of freedom. We used our own judgment but we knew what the ideals were and what kind of things they would want us to do and didn't want, but they never told us outright. Looking back on it, it's kind of a mystery to me. We were never overtly praised but there was a subtle awareness on our part that our parents were appreciative. There was perhaps a tendency to play down success a little bit to prevent becoming cocky or too self-satisfied. I don't think I ever became rebellious; I suspect the small local environment was restrictive as much as the family."

The family life of one subject was apparently relatively placid. His mother was a gifted person who sang and composed quite spontaneously and was very social and warm. About the only open pressure put on the children was an attempt to make them practice two hours a day but this was given up, largely apparently because of the fight put up by an older brother. "My father was a man who values independence and self-reliance and he is anxious to see his sons have it." To this end he gave the son a good deal of support, both personal and financial, and there is no evidence of any serious problems arising. Their present relations seem closer than any others in this group.

Another reported differences in parental interests. "He was a very austere and intellectual person and she was a warm, affectionate one. I suspect there was some unhappiness there." I have no record of any particular difficulties between him and his father, who certainly gave him considerable vocational encouragement, but it may be that his early enlistment served to avoid this. "I was very innocent and protected when I went in, I was just 18. At first I was kicked around. I would write in a diary and read books and everybody thought that was very funny. At first it bothered me the way they carried on about it, but after I took up boxing and made a place for myself they left me alone. I realize now I had been looked after too much, that there has been this pattern of extended protection not only because I was an only child but Mother had her own ideas and she was anxious to bring me up in the right way. Father, I realize now, was pretty much under my mother's thumb. It's undoubtedly true I had a very close attachment to her and revolted, but it didn't come to any actual crisis, there was a general withdrawal. I was away for so long."

"Dad let Mother be the boss in spiritual and moral matters but he was a titanic person in his own way. There was a division of labor that seemed authentic and all of us children felt that our parents were adequate. There was no rejection and unstructuredness about it. There was a time when I hated my father, at least I hated his sort of crudeness. I had quite a rebellion thinking he was a hard man and not understanding. I think I had a normal amount of adolescent separateness and rejection. But I was away from home and I didn't have the friction and I was making my own way."

(In the rest a variety of problems appear, some of them of great severity.)

"I didn't see a great deal of my father. I guess I took Mother's side. I liked him when he was nice to me. I was always a little afraid and upset by him. I recognize now it wasn't so much by him but the situation involving mother. Later on it looked as though my aunt and father were battling to see who would be head of the family. I'm afraid I adopted a retreating attitude."

"I think family discipline was very strict as I look back on it. I think father was a very strict person and I'm sure I had very strong fears of his censure. I went through a religious period about 13 or 14, it was very intense, then I rejected my father and God all in one fell swoop. It stuck for God but not for Father. I don't believe I ever rejected him or his authority. I never fought the battle out with Father. I never tried to argue with him. I just shut up and hoped he would not ask questions. I think all I have done has been kept this side of a line to avoid his displeasure but I didn't feel any relief at his death a few years ago."

The history of one subject is of revolt against one authority after another, although it is not altogether clear just how his family figure in this. "Mother was pretty much under my father's thumb. I was generally fairly antagonistic to her, she was always trying to stop me from trying to do things. There was always a lot of argument about drinking. Mother didn't approve but father invited us to drink with him in the cellar."

"In my opinion a great deal of the pattern of my life is to be understood as a reaction formation against my father." The family situation was socially and personally a complicated one, and this subject is probably correct in his interpretation. The forms of his revolt were also complicated, but before he went to college there was open enough friction between him and his father that a family friend interfered.

For another subject relations with his father were of considerable importance. "I had great respect for him. I thought he was one of the smartest and most successful people I had ever known. I also had very little affection I can recognize. He instilled in me the feeling that I was bad, so I grew up with the impression that I was the sort of person people did not like to have around, and if I walked down the street and one of my companions was across the street and didn't happen to see me and didn't speak to me, I just took it as further evidence of the fact that people didn't like me. Mother never gave in to me but at the same time she was much gentler and more understanding, and even if she didn't understand she would go along with me."

"Father seemed to me to typify a rather clear culture pattern that existed through the Middle West, the pattern of the good moral family, hard-working and making their own way. He is a tense man, rather stern, affectionate, but terribly pressed with ambition. He was very punitive about aggressive behavior and very rigid about sex and probably a very anxious person and insecure in his social relations. My mother is a very neurotic woman, hypochondriacal and hysterical. There was nothing stern about her, she gave in to us. Her methods of discipline were via the channel of idealizing and I think my father used that, too. 'You don't love me' sort of thing. They had a device for putting serious arguments about things in a semi-joking way. I don't recall seeing any direct anger between them ever, and I think that has played an important role in my development. Then my brother got all the attention and I had a rough time but I guess I gave him a rougher one. I revolted violently in adolescence but as a rebellion it wasn't conscious at all. I never phrased it as that."

Another had a history of serious family problems from the start. His maternal grandmother lived with them and set herself to make trouble between the parents and particularly between him and his father. "My father was exceedingly hostile where I was concerned and exceedingly unpredictable. I could never tell, when he came into the room, whether he was going to be nice to me or knock me down. My mother was very intelligent but sweet and negative, and never interfered on my behalf as far as I knew as a child. She simply stood aside, and of course while I was fond of her I had little respect for her. So the situation was I had a feeling from a very early age that I had to be completely dependent on myself, which resulted in a diminution of emotional affect. I was never close to anybody as a child. The thing that saved me from more serious psychological involvement was that I hated my father overtly."

It is clear that patterns involving overprotection and firm, if not overt, control are very common in the group. They are commoner among psychologists than among anthropologists among whom there was more overprotection and more open hostility. Over half of this group reacted with more rebelliousness than is generally usual, and of these a number are still angry or rejecting or disrespectful of one or both parents.

Comparison with Other Scientists

The data on intrafamily relations are more complete than for the other groups, partly because of the fact that these groups are professionally more aware of the possible significance of such relations and are generally freer in such discussions. But there is additional, if inferential, evidence from the TAT protocols, and I think there is no doubt that the groups do differ in these respects. Both the physicists and the biologists early developed ways of life which involved very much less of personal interaction, and neither group shows anything like the extent of rebelliousness and family difficulty that the psychologists and anthropologists show.

There are also many more in the other groups who were isolates as children, or who had only one or two close friends, and the age of beginning heterosexual interests is very different. Among the biologists and physicists it is rare for there to have been any extensive dating in high school or early college. Half of the social scientists began dating in high school and dated happily and extensively from then on. Only four of them did very little or no dating until they were through college. Two of the psychologists apparently never dated any girl but the one each married (rather late in life), and have never had much social life since. These are very atypical for the group. Although a number of the social scientists, particularly those from self-engrossed homes, speak of being shy for a time, shyness was rarely the serious problem that it was with many of the biologists and physicists.

In the earlier monographs some space was devoted to consideration of present recreational interests. Inquiry was made about these in the social scientist group also, but as they differ from the others only in their generally greater interest in social life, which is a continuation of the patterns reported above, further details are not given.

VI. RELIGION

There are two Jewish families in the social scientist group, one devout and the other non-practicing. The others are all Protestant families, with most of the major churches and several of the smaller groups represented. The parents usually attended church, but frequently for social reasons. Only two of the subjects ever go to church now and one does not do so for religious reasons.

About half of the parents of psychologists were personally uninterested in religious matters (as demonstrated in church attendance), but only two families made a point of not attending church. Nevertheless it was customary to send the children to Sunday School and all of the Protestant children did go for at least a time, even the son of freethinkers. Seven of the subjects stopped going fairly early, and while two encountered some family opposition in withdrawing there was no personal crisis or conflict over religion. Several quotations are illustrative.

> "Religion was one thing we were saved. Mother went to church when she was asked to sing and usually went once more for every time she was asked. Father would go occasionally with her because it was the proper thing to do. And I went to Sunday School and had the best necking of my life, but I don't think I've ever been in a church since."

> "Both parents were church members but Dad never went. Mother was an Episcopalian and still goes sporadically. I was encouraged but not forced to go to Sunday School and I went fairly regularly. At 10 I shifted on the basis of the basketball court. I never joined a church and I never wanted to. After contact with a psychology professor and reading *This Believing World* which I got hold of somehow, I felt this isn't for me. Religion has never been a source of conflict for me in any way."

Five others were quite active in various ways, in young peoples' societies, in the YMCA, in teaching Sunday School, and continued their interest through college or beyond. Three of these were professionally interested, one actually serving as a missionary for some time and two studying religious education. For example:

> "There was very much church influence. My parents were very religious and belonged to a group where the religion is quite emotional. I couldn't understand what was going on and couldn't appreciate it, I was bewildered and sometimes frightened. Church attendance was required and Sunday was very strictly observed in our family. At college I attended church services and Sunday School and was a leader in Christian Endeavor for a year or so. But I had no serious crisis. I had never taken the funda-

mentalist point of view, I had always questioned it and tried to analyze religion as such."

One subject went through an intense religious phase in the early teens of a rather extreme sort. This was not closely connected with any organized church activities but a very personal matter. "I wrote a Bible and then carefully folded it up and hid it, because I was very much afraid that my brother would find it. I had several miracles happen to me which I can't explain even now. I wrote them all out in Biblical prose. I certainly had some kind of sense of history and destiny. I suddenly threw it all over but there was no struggle."

The one psychologist who now attends church regularly, for religious reasons, states, "I find more wisdom concerning mind, values, conduct, and therapy in Christian doctrine than I do in our beloved but still puerile science of psychology."

Among the anthropologists the picture is similar. One of them who came from a Jewish family had a little religious schooling, but it was not important to his parents or to him. One was the son of parents who had a family tradition of agnosticism on both sides, and religion never was a concern to him. The other six were all sent to Sunday School, although the parents of two were personally uninterested. Four of these soon dropped out, usually because of boredom, but the other two retained interest for some time, but have no church affiliations now.

COMPARISON WITH OTHER SCIENTISTS

Of the 64 scientists studied altogether, whose religious backgrounds were not known when they were selected, none came from Catholic families. Five came from Jewish homes, and all of the rest had Protestant backgrounds. These include two Mormons and two Quakers.

Among all of them the picture is much the same. Most went to Sunday School; very few now have any church connections. Two biologists are very active in church work; another contributes to a church but does not attend. Among the physical scientists none is personally active in any church, although there are five who have maintained some church connections, usually to please their wives. Among the social scientists, one is still personally interested in church, one goes ocassionally, but not for religious reasons, and one pays dues but never attends. A few of them are militantly agnostic, but for the most part they are just not interested.

VII. THE VERBAL-SPATIAL-MATHEMATICAL TEST

This test (VSM) was compiled for the study by the Educational Testing Service. The verbal test contains 79 items in two sections, in each the task being the selection of antonyms. Time limit was 15 minutes for the two sections. The spatial test comprised 24 items, with a time limit of 20 minutes. The task was to select from four stimuli, the two views of the same figure. The mathematical test comprised 39 items, of mathematical reasoning. Time limit was 30 minutes. Examples are given in the earlier publications (20, 23).

Results are given in terms of sigma scores, in Table 6. The scores for the psychologists are based on the distribution of psychologists, for the anthropologists on the distribution of anthropologists. Two of the anthropologists declined to attempt the spatial and mathematical sections. The difference in means between psychologists and anthropologists is not significant for the verbal test, but t for the difference between means for the spatial test is 6.88 and $p < .01$; for the mathematical test, t is 6.68 and $p < .01$.

All but one of the experimental psy-biologists and physicists it is rare for there to have been any extensive dating in high school or early college. Half of the social scientists began dating in high school and dated happily and extensively from then on. Only four of them did very little or no dating until they were through college. Two of the psychologists apparently never dated any girl but the one each married (rather late in life), and have never had much social life since. These are very atypical for the group. Although a number of the social scientists, particularly those from self-engrossed homes, speak of being shy for a time, shyness was rarely the serious problem that it was with many of the biologists and physicists.

In the earlier monographs some space was devoted to consideration of present recreational interests. Inquiry was made about these in the social scientist group also, but as they differ from the others only in their generally greater interest in social life, which is a continuation of the patterns reported above, further details are not given.

VI. RELIGION

There are two Jewish families in the social scientist group, one devout and the other non-practicing. The others are all Protestant families, with most of the major churches and several of the smaller groups represented. The parents usually attended church, but frequently for social reasons. Only two of the subjects ever go to church now and one does not do so for religious reasons.

About half of the parents of psychologists were personally uninterested in religious matters (as demonstrated in church attendance), but only two families made a point of not attending church. Nevertheless it was customary to send the children to Sunday School and all of the Protestant children did go for at least a time, even the son of freethinkers. Seven of the subjects stopped going fairly early, and while two encountered some family opposition in withdrawing there was no personal crisis or conflict over religion. Several quotations are illustrative.

> "Religion was one thing we were saved. Mother went to church when she was asked to sing and usually went once more for every time she was asked. Father would go occasionally with her because it was the proper thing to do. And I went to Sunday School and had the best necking of my life, but I don't think I've ever been in a church since."

> "Both parents were church members but Dad never went. Mother was an Episcopalian and still goes sporadically. I was encouraged but not forced to go to Sunday School and I went fairly regularly. At 10 I shifted on the basis of the basketball court. I never joined a church and I never wanted to. After contact with a psychology professor and reading *This Believing World* which I got hold of somehow, I felt this isn't for me. Religion has never been a source of conflict for me in any way."

Five others were quite active in various ways, in young peoples' societies, in the YMCA, in teaching Sunday School, and continued their interest through college or beyond. Three of these were professionally interested, one actually serving as a missionary for some time and two studying religious education. For example:

> "There was very much church influence. My parents were very religious and belonged to a group where the religion is quite emotional. I couldn't understand what was going on and couldn't appreciate it, I was bewildered and sometimes frightened. Church attendance was required and Sunday was very strictly observed in our family. At college I attended church services and Sunday School and was a leader in Christian Endeavor for a year or so. But I had no serious crisis. I had never taken the funda-

mentalist point of view, I had always questioned it and tried to analyze religion as such."

One subject went through an intense religious phase in the early teens of a rather extreme sort. This was not closely connected with any organized church activities but a very personal matter. "I wrote a Bible and then carefully folded it up and hid it, because I was very much afraid that my brother would find it. I had several miracles happen to me which I can't explain even now. I wrote them all out in Biblical prose. I certainly had some kind of sense of history and destiny. I suddenly threw it all over but there was no struggle."

The one psychologist who now attends church regularly, for religious reasons, states, "I find more wisdom concerning mind, values, conduct, and therapy in Christian doctrine than I do in our beloved but still puerile science of psychology."

Among the anthropologists the picture is similar. One of them who came from a Jewish family had a little religious schooling, but it was not important to his parents or to him. One was the son of parents who had a family tradition of agnosticism on both sides, and religion never was a concern to him. The other six were all sent to Sunday School, although the parents of two were personally uninterested. Four of these soon dropped out, usually because of boredom, but the other two retained interest for some time, but have no church affiliations now.

COMPARISON WITH OTHER SCIENTISTS

Of the 64 scientists studied altogether, whose religious backgrounds were not known when they were selected, none came from Catholic families. Five came from Jewish homes, and all of the rest had Protestant backgrounds. These include two Mormons and two Quakers.

Among all of them the picture is much the same. Most went to Sunday School; very few now have any church connections. Two biologists are very active in church work; another contributes to a church but does not attend. Among the physical scientists none is personally active in any church, although there are five who have maintained some church connections, usually to please their wives. Among the social scientists, one is still personally interested in church, one goes ocassionally, but not for religious reasons, and one pays dues but never attends. A few of them are militantly agnostic, but for the most part they are just not interested.

VII. THE VERBAL-SPATIAL-MATHEMATICAL TEST

This test (VSM) was compiled for the study by the Educational Testing Service. The verbal test contains 79 items in two sections, in each the task being the selection of antonyms. Time limit was 15 minutes for the two sections. The spatial test comprised 24 items, with a time limit of 20 minutes. The task was to select from four stimuli, the two views of the same figure. The mathematical test comprised 39 items, of mathematical reasoning. Time limit was 30 minutes. Examples are given in the earlier publications (20, 23).

Results are given in terms of sigma scores, in Table 6. The scores for the psychologists are based on the distribution of psychologists, for the anthropologists on the distribution of anthropologists. Two of the anthropologists declined to attempt the spatial and mathematical sections. The difference in means between psychologists and anthropologists is not significant for the verbal test, but t for the difference between means for the spatial test is 6.88 and $p < .01$; for the mathematical test, t is 6.68 and $p < .01$.

All but one of the experimental psy-

TABLE 6

The Verbal-Spatial-Mathematical Test Sigma Scores

Subject	Psychologists			Subject	Anthropologists		
	V	S	M		V	S	M
Ps 1	+0.32	−0.55	+1.10	An 1	+0.57	−0.04	−0.67
Ps 2	−0.21	−0.55	+0.80	An 2	+0.77	+0.84	−0.06
Ps 3	+0.17	−0.78	+1.25	An 3	+1.22	−1.15	+1.15
Ps 4	+0.24	+0.41	−0.53	An 4	+0.26	−0.49	+0.85
Ps 5	−1.83	−0.07	−0.90	An 5	−0.14	−0.71	−1.58
Ps 6	+0.39	−1.26	−0.90	An 6	+0.55	+1.51	+0.24
Ps 7	+0.14	+1.36	−0.08	An 7	−2.03		
Ps 8	−2.05	−0.07	−0.68	An 8	−0.12		
Ps 9	+0.54	+0.89	+1.54				
Ps 10	+1.13	+1.84	+1.69				
Ps 11	+0.17	+0.65	−1.12				
Ps 12	+0.32	−1.02	−0.53				
Ps 13	+0.84	−1.50	−0.90				
Ps 14	+0.61	+0.65	−0.53				
Mean	57.7±3.6	11.3±1.1	15.6±1.8		61.1±3.2	8.2±1.8	9.2±1.4
SD	13.5±2.5	4.2±0.8	6.8±1.3		8.9±2.2	4.5±1.3	3.3±0.9

chologists has a higher sigma score for either spatial or mathematical than for verbal. Two of the others have their highest scores on the spatial test and two on the verbal, the difference in one instance being very slight. No tendencies are evident among the anthropologists.

TABLE 7

Comparison with Other Scientists on the Verbal-Spatial-Mathematical Test

Test	Biologists ($N=19$)	Physicists ($N=18$)	Social Scientists ($N=22$)
Verbal			
N right, range	28–73	8–75	23–73
Mean	56.6±2.8	57.3±4.1	59.0±4.2
Spatial			
N right, range	3–20	3–22	3–19
Mean	9.4±1.0	13.0±1.2	10.4±0.9*
Mathematical			
N right, range	6–27		4–27
Mean	16.8±1.4		13.7±1.5*

* $N=20$

The intercorrelations for the social scientists on this test are: verbal-spatial, +.18; verbal-mathematical +.27; spatial-mathematical +.36. None is significant.

Comparison with Other Scientists

Table 7 presents the material for comparison with the other groups studied. The mathematical test was not difficult enough for the physicists. Differences between the means of the different groups are small and not significant. It should be noted that there is a large difference between the subgroups of physicists on the verbal test, the experimentalists averaging 46.6 and the theorists 64.2. On the spatial test, their averages are 11.7 and 13.8 respectively. If comparison is made by analysis of variance for five groups, experimental physicists, theoretical physicists, biologists, psychologists, anthropologists, F approaches the 5% level, even in these small groups.

Intercorrelations for the total group are given in Table 8. Correlation with age is significant only for the spatial test, with a $p < .01$. Of test intercorrelations, only the verbal-spatial reaches this level. It is clear from descriptions by the subjects that the spatial test can be done in various ways, and in part by verbal reasoning. The total distribution for the verbal test is strongly positively skewed; the spatial test distribution is platykurtic and the mathematical is bimodal.

TABLE 8

INTER-TEST CORRELATIONS, AND CORRELATIONS WITH AGE: TOTAL GROUP

Variable	Age	Verbal	Spatial
Verbal			
r	−.11		
N	59		
Spatial			
r	−.40	+.33	
N	57	57	
Mathematical			
r	+.00	+.14	+.21
N	39	39	39

COMPARISON WITH GRADUATE STUDENTS

Through the courtesy of Dr. Irving Lorge of the Institute of Psychological Research at Teachers College, Columbia University, the VSM was given to all candidates for doctoral degrees in either philosophy or education in February, 1951. At the same time a number of other tests were administered to these 174 students. These data are presented in Table 9. They throw light on the nature of the test used in this study and make it possible to transmute raw scores on the VSM to scores on the Scholastic Aptitude Test. The two sections of the verbal test were handled separately in the correlations, but were combined in computing equivalents on the Scholastic Aptitude Test.

From the correlation matrix it would seem reasonably clear that the verbal test used here is a measure of the same function tested by Lorge's total verbal measure. Both the spatial and mathematical tests seem largely independent of the verbal test and of each other, the mathematical less so.

Transformation of scores on the VSM to Scholastic Aptitude Test scores is given in Table 10. The SAT equivalents of various percentile scores are given for each subtest. The normal mean and standard deviations for this test for applicants to college as undergraduates are 500 and 100. For doctoral candidates at Teachers College these figures are 570 and 130. These figures are not available for any other school. VSM equivalents for this average SAT score are 32 on the verbal test, 11 on the spatial test, and 8 on the mathematical. Five of the scientists are below the mean on the verbal, 29 on the spatial (but it must be remembered that this test correlates −.40 with

TABLE 9

CORRELATIONS OF VSM WITH OTHER TESTS

($N = 174$)

	Verbal I	Verbal II	Mathematical III	Spatial IV	Mean	*SD*
Completion	.716	.621	.204	.369	510.8	140.6
Vocabulary	.800	.715	.056	.254	555.0	128.7
Eng. Place Voc.	.823	.683	.178	.289	554.8	147.9
Reading	.652	.546	.322	.409	504.9	167.6
Total Verbal	.796	.681	.268	.416	531.2	132.6
Information	.318	.177	.287	.299	587.4	115.7
Arithmetic	.382	.278	.344	.764	593.5	122.1
English Usage	.662	.506	.283	.351	561.7	134.0
Reading Comprehension	.540	.399	.284	.363	532.5	157.6
Reading Speed	.423	.331	.256	.276	534.5	140.4
VSM, Verbal I		.795	.060	.299	22.0	11.5
VSM, Verbal II			−.091	.211	9.4	6.4
VSM, Mathematical				.378	8.0	5.1
VSM, Spatial					10.7	6.6

TABLE 10

SCHOLASTIC APTITUDE TEST EQUIVALENTS FOR VSM RAW SCORES

Percentile	Verbal I and II	SAT Equivalent	Spatial	SAT Equivalent	Math.	SAT Equivalent
100	75	892	22	784	27	1042
75	67	833	15	651	22	918
50	61	788	10	556	13	694
25	52	722	7	499	10	619
0	8	395	3	423	4	470

age) and 3 on the mathematical.

It is clear that the average ability of the scientists is very great. This is not surprising. On the other hand, it is surprising, and a matter of very considerable importance, that there are among the scientists a number who are not facile at the types of tasks presented by the VSM, but who have been able to make contributions of great value to society.

VIII. THE THEMATIC APPERCEPTION TEST

This test, devised by Murray and his associates (12), is a technique for personality analysis. It consists of a set of pictures, the task of the subject being to tell a story about each picture, including the events leading up to the moment pictured, what is going on, what the characters are thinking and feeling, and what is going to happen in the future. Only 9 of the usual 20 cards in the series were used in this research. They are cards 1, 2, 4, 6, 7, 10, 13, 15, and 11, presented in that order, from the male series of 1943. The TAT was always given after the Rorschach.

These social scientists have little knowledge of the TAT and there is no one among them who is expert in its scoring and interpretation. One of the psychologists and one of the anthropologists do have a considerable acquaintance with TAT theory, and the anthropologist has given, but not interpreted, the test. Seven of the psychologists and one anthropologist have read about the test or heard discussions of it, but have no technical knowledge, and the others know nothing about the test, although a few have seen the cards around.

The TAT is difficult to handle as a research instrument, since the scoring is not well codified, but it supplements the interview material elegantly for individual analysis. As in previous studies (16, 22), I have again followed Wyatt. The basic data are presented in Tables 11, 12, 13, and 14. In all of the tables the entries in the columns are the numbers designating particular pictures. Numbers are placed in parentheses when the column heading applies to part of the story but not all of it (e.g., if the story has an unusual twist, but is not entirely unusual).

In Table 11 the first series of entries refers to the relative amounts of narration (S) and description (D). Purely descriptive responses, usually a form of noncompliance, are uncommon.

Perceptual distortions were so rare that they have not been tabled. The columns under the heading Perception refer to details disregarded or given considerable prominence.

TABLE 11

THEMATIC APPERCEPTION TEST (TAT): SUMMARY, PART I

(Numbers in the table entries refer to TAT cards. For explanation of column headings, see text.)

Psychologists	Story					Perception			Time Trend			
	Refuse	S	SD	DS	D	Omit	Descrip. Detail	Deviation	Past, Present, Future	Past, Present	Present, Future	Past Only
Ps 1		all				1, 2, 4, 11		(2), 15	1, 4, 6, 7, 11	2, 10	15	13
Ps 2		2, 4, 6, 7, 13, 15	1, 11		10	1	2	(4)	2, 7		1, 4	6, 10, 13, 11, 15
Ps 3		all (10 stories)				1, 2		(4), 15	4, 11	4, 11		1, 2, 6b, 7, 10, 13
Ps 4		4, 6, 7, 10, 13, 15, 11	2	1			1	(1), 15	4, 10, 13	15	1, 2, 6, 11	7
Ps 5		4, 6, 7, 10, 13abc, 15	1, 2ab	11		11	1	13b, (15)	1, 13b	10	2b, 6, 13c	2a, 4, 7, 13a, 15, 11
Ps 6		all						1, (4), 7, 13 15, 11	all but 10			10
Ps 7	2, 10, 11, 13	1, 4, 6		7	15		4					1, 4, 6, 7, 15
Ps 8		all but 11	11			1		6, 7, 10, (15), 11	1, 2, 4		6, 7, 10, 13, 15	11
Ps 9		6, 7, 15	1, 2, 4, 11	13	10	1, 4, 13		(1), 4, (6)	1, 2, 4, 6, 7, 13			10, 15, 11
Ps 10	15	1, 4, 6, 7, 13	2, 11		10		1, 2, 4, 13, 11	(1), (7)	2, 6	13	1, 4	7, 10, 11
Ps 11		all				1, 4, 13		(1), (6), (7), (10), (13)	1, 2, 4, 6, 7, 13, 15, 11	10		
Ps 12		all but 2	2			1		(6), (10), (13), 15, 11	all			
Ps 13		all				1	2	(6), (10)	2, 6		1, 4, 10, 13, 11	7, 15
Ps 14		all but 2	2			1	11	11	1, 4		6, 13, 15, 11	2, 7, 10
Anthropologists												
An 1		7a, 13, 15	6abc, 1, 2, 4, 7b, 10, 11			1	all	15	1, 2, 4, 7b, 10, 13, 15, 11	6a, 7a		6bc
An 2		all but 2			2	1		(4), (6), (7), (10), 13, 11	6, 7, 11		1, 10, 13, 15	2, 4
An 3		all but 1b	1b			1a		10, 15, 11	4, 10	1a, 2, 6, 7, 13	11	1b, 13
An 4	11	all (8 stories)				1		1, 2, 7, 10, 15, (4), (6), (13)	1, 10, 13	7	2, 4	6, 15
An 5		4, 6, 7, 10 13, 15	2, 11		1			2	7, 13		4, 6	1, 2, 10, 15, 11
An 6		2, 6, 13	1, 7, 11	4, 10	15	2, 4		(13)	2, 13	1, 6, 7, 10, 11		4, 15
An 7		all but 2	2					(2), (6), 10			1, 13, 11	2, 4, 6, 7, 10, 13
An 8		1, 4, 6, 7, 13, 11	2	15	10	1		13			1, 2, 4, 7, 11	6, 10, 13, 15

TABLE 12

THEMATIC APPERCEPTION TEST (TAT): SUMMARY, PART II

(Numbers in the table entries refer to TAT cards. For explanation of column headings, see text.)

Psychologists	Outcome					Certainty of Outcome			Personal Reference or Opinion
	None Possible	Success	Defeat	Unsolved Tension	None or ? Given	Certain	Probable	Possible	
Ps 1		4, 11	1, 7, 15	6	2, 10, 13	1, 4, 7, 15, 11	6		
Ps 2		1, 2, 4			6, 7, 10, 13, 15, 11	2	1, 4		15
Ps 3		6a, 15		13	1, 2, 4, 6, 7, 10, 11	15	6a	13	6b
Ps 4		2, 6, 10	1, 4	11	7, 13, 15	1, 4, 6, 10, 11	2		
Ps 5	15	1, 6	2ab	13abc	4, 7, 10, 11		1, 2, 6, 13abc		
Ps 6		1, 2, 6, 13	4, 7, 15, 11		10	1, 2, 4, 6, 13, 15, 11	7		10
Ps 7					1, 4, 6, 7, 15				13
Ps 8		1, 2, 4, 6, 7	10, 13	15	11	2, 4, 6, 10, 13, 15	1, 7		
Ps 9		1, 2, 6, 7	13	4	10, 15, 11	2	1, 4, 6, 13	7	1, 2
Ps 10		1, 2	6	4	7, 10, 13, 11	2	1, 6	4	
Ps 11		10, 11	1, 2, 13, 15	4, 6	7	10, 13, 15, 11	1, 2, 4, 6		7
Ps 12		4, 7, 10, 13, 11	2, 6, 15	1		2, 4, 6, 10, 13, 15, 11	1, 7		
Ps 13		6, 10	2	1, 4, 13	7, 15, 11	2, 4, 10	6, 13	1	
Ps 14		1, 2, 4, 6, 10, 13, 15, 11			7	6	all but 6, 7		(11)
Anthropologists									
An 1		1, 2, 7b, 10, 11			6abc, 7a, 13, 15	1, 7b, 10, 11	2, 4		
An 2		1, 6, 7, 13, 11	10		2, 4, 15	1, 6, 7, 10, 13, 11			
An 3		4, 10	11		1ab, 2, 6, 7, 13, 15	4, 10, 11			
An 4		2, 4, 10	13		1, 6, 7, 15	2, 4, 10, 13			1, 2
An 5		4, 6, 7		13	1, 2, 10, 15, 11	4, 6, 13	7		
An 6		10	2, 6, 13, 11	1, 7	4, 15	2, 6, 10, 13	7, 11	1	
An 7			13	1, 11	2, 4, 6, 7, 10, 15	13, 11	1		
An 8		2, 4, 7		1	6, 10, 13, 15, 11	1, 7	2	4	

Deviation refers to unusual stories. Cards 15, 11, and 10 are the ones most often entered under this heading.

Time trend is given in the last section of the table. If a subject omitted part of the time span he was asked about it but not pressed for it. The psychologists followed instructions to give a full time span significantly more frequently than the anthropologists did (chi square was 5.57, $p < .02$), with more of the anthropologists omitting the future. About half in each group gave no past.

In Table 12, outcome of the stories

TABLE 13—THEMATIC APPERCEPTION TEST (TAT): SUMMARY, PART III

(Numbers in the table entries refer to TAT cards. For explanation of column headings, see text.)

	Level						Tone								
Psychologists	Concrete Factual	Endo-psychic	Symbolic	Mythical	Make-believe	Conditional	Indifferent Detached Con-templat.	Cheerful Serene	Unhappy	Tense	Anxious	Morbid	Aggressive	Melo-dramatic	Sardonic
Ps 1	1, 4, 7, 13, 11	2, 6, 10, 15	(11)				11	4	1, 2, 6, 7		10		13	15	
Ps 2	1, 2, 4, 6, 7, 10	15			11	13		1, 2, 10	4, 15	6	7, 13		11		
Ps 3		all including 6b							6ab, 13, 15	7, 10	1, 2, 4			11	
Ps 4	all but 11		(2), (4), (13), (15), 11					10?	1, 4, 6		2, 7, 15, 11		13		
Ps 5	11	all but 11							2ab, 6, 10, 13ab, 15	1, 4, 7, 11	13c				
Ps 6	all but 10			10				(1, 4, 6) 2, 13	1, 4, 6, 7, 10				15		11
Ps 7	4, 6, 7	1			Dr 15					1, 6, 7	15		4		
Ps 8	all but 15	15							6, 10	1	2, 4, 7, 15		13, 11		
Ps 9	1, 2, 4, 6, 7, 10	15	(1), (6)	11	13	(13)	2, 10	1		6	7		4, 13		15, 11
Ps 10	1, 2, 4, 6, 7	10, 13	(4)	11				2, 10	6, 7	1	4, 13	(7)	11		
Ps 11	11	all but 11						10	1, 2, 6, 13, 15	4		(13), (15)	7	11	
Ps 12	1, 4, 6, 7, 10, 13, 15		2	11				(7), (10), (13)	2, 4, 6, 7, 10, 13, 15	1			11		
Ps 13	6, 11	1, 2, 4, 7, 10, 13, 15							10, 15	1, 2, 4, 6, 7, 13			11		
Ps 14	2	1, 4, 6, 7, 10, 15, 11			13			1, 2	15	7, 10, 11	4.6		13		
Anthropologists															
An 1	7a, 11	1, 2, 4, 6abc, 7b, 10, 13, 15						1, 2, 10	6abc, 7a, b, 13			15		4, 11	
An 2	4, 6, 7, 10	13, 15	2		11		2	1	6	4, 7			10, 11		13, 15
An 3	4, 6, 7	1ab, 2, 10, 13	(11)	15, 11			1a	1b, 2	4, 6, 7			15		10, 11, 13	
An 4	1, 2, 6, 10, 13	7, 15			4									4, 13	1, 2, 6, 7, 10, 15
An 5	2, 7, 13	1, 4, 6, 10	15	11		(6)	1		2, 15	10, 11	4, 6, 7	13			
An 6	7, 13, 11	1, 2, 6, 10	15		4				1, 2, 6, 13, 15, 11	7, 10				4	
An 7	2, 4, 6, 7, 13, 15, 11	1			10			2	15	1, 4, 6, 7, 10			13, 11		
An 8	all but 15	15						10	6	1, 2, 7, 15			4, 13, 11		

and the certainty with which the outcome is stated are recorded. Half of the stories of the anthropologists and a third of those of the psychologists do not give an outcome, or give an unclear one. In both groups a successful outcome is predicted about one third of the time. Defeat and unsolved tension make up the remainder in about equal amounts. But of the stories to which an outcome is given, the anthropologists have a higher degree of certainty in their statements. This may just mean a balancing of their greater caution in giving an outcome—they give it only when they are sure. Stories with personal references occur rarely.

In Table 13, level of the stories is recorded, as concrete-factual (C-F); endopsychic (E-P); symbolic (Sym); mythical or past (Myth); make-believe (M-B); and conditional (Co.) There are no differences between psychologists and anthropologists in these respects. Stories that are dominantly concrete-factual are a little commoner in both groups than are stories concerned primarily with what the characters are thinking and feeling. The other levels are rarely used.

General tone of the stories is also recorded in Table 13. The first column in this section includes stories scored as indifferent, detached, or contemplative; the second column, those labelled cheerful or serene. The other columns are unhappy, tense, anxious, morbid, aggressive, melodramatic and sardonic. Tones labelled unhappy, tense, and anxious are chiefly differences of degree, and incidence of stories in these groups is considerably higher for the psychologists than for the anthropologists, who have more recourse to the melodramatic and sardonic.

For the social scientists quality of stories has not been tabled, since they are rarely scored anything but literate.

In Table 14 are given the types of personal relations described for the characters in the stories, and the nature of the process attributed to the situations. For both groups, formal relations (e.g., father, son; husband, wife) are a little commoner than emotional ones (e.g., lovers, friends). There are some differences in presses, which do not quite reach significance at the 5% level (5 × 2 chi square table, with chi square 9.85). The anthropologists give more unfriendly and the psychologists more internal presses.

There are a few themes that seem to recur with considerable frequency, the most evident being one of general helplessness in the face of severe problems. This is sometimes very general and sometimes seems to be limited to the male figures in the stories in contrast to the female figures. There is also a considerable feeling of dependence on parent figures, and while there are some stories of successful rebellion without serious guilt feelings, there are more stories of characters who rebelled only with guilt and general unhappiness. At the same time there are a few who seem to have a strong sense of responsibility with regard to human relations.

The details on heterosexual relations are usually fairly full. This group of social scientists is not particularly conventional in its approach but is definitely much concerned with interpersonal relations and finds it relatively easy to verbalize them.

There are some among them who find contemplation of death a serious problem. In two instances this concern with death may be a major factor in their professional activities.

TABLE 14

THEMATIC APPERCEPTION TEST (TAT): SUMMARY, PART IV

(Numbers in the table entries refer to TAT cards. For meaning of column headings, see text.)

Psychologists	Personal Relations				Presses				
	Formal	Emotional	Not Stated	None	Friendly	Unfriendly	Impersonal	Internal	?
Ps 1	1, 6	2, 4, 7, 10, 13, 11		15	1, 4, 6, 7	10	2, 7, 11	15	
Ps 2	2, 4, 6, 7	10, 15	13	1, 11	4, 7, 10	11	15		1, 2, 6, 13
Ps 3	2, 4, 6, 7, 13, 15	10, 11		1	2	6b	1, 4, 7, 13, 15	6a	10, 11
Ps 4	2, 6, 7, 10, 15	4	13	1, 11	10			all but 10	
Ps 5	1, 2a, 4, 6, 7, 10, 13a, 15	2b, 13bc, 11			1, 4, 7	2ab, 6, 11	10, 13ab, 15	13c	
Ps 6	2, 4, 6	7, 10, 13, 15, 11		1	11	6, 7, 15	1, 2	4, 7, 13	
Ps 7	4, 6, 7			1, 15	6	4			1, 7, 15
Ps 8	1	2, 4, 6, 7, 10, 13		15, 11	1, 7	2, 13	6, 10, 15, 11		4
Ps 9	1, 2, 6, 7	4, 15		10, 13, 11	1, 10	13	2, 15	4, 6	7, 11
Ps 10	2, 6	4, 7, 13	10	1, 11	2, 4, 10	7	6	1, 13	11
Ps 11	1, 2, 10	4, 6, 7, 13, 11	15		4	7	6, 10, 11	2, 13, 15	1
Ps 12	1, 2, 6, 10	4, 7, 13		15, 11		15	2, 4, 6, 7, 10, 11		1, 13
Ps 13	2, 4, 6, 7, 10	13		1, 15, 11	7	11	6, 10, 15	1, 2, 4, 13	
Ps 14	2, 6, 7, 15	4, 13	10	1, 11	15	4, 6, 13, 11	10	1, 2	7
Anthropologists									
An 1	1, 2, 6c, 7a, 10	4, 6ab, 7b, 13, 15, 11			1, 4, 10, 13	2, 15	6abc, 7ab, 11		
An 2	6, 7, 10, 15	1, 4, 13, 11		2	1, 2, 7	4, 10, 11	15	6, 10	
An 3	1a, 4, 6, 7	2, 10, 13, 15		1b, 11	1a, 10, 13	11	7, 15	1b, 2, 4, 6	
An 4	2, 4, 6, 10, 13	1, 7		15	6, 7	1, 4, 10, 13	2		15
An 5	6, 7, 10	2, 4, 13		1, 15, 11	7	2, 10, 13, 15	6	4	1, 11
An 6	1, 2, 6, 10	4, 7, 13		15, 11		1, 2	4, 6, 10, 13, 11		7, 15
An 7	1, 2, 6, 7, 15	4, 10, 13		11	1, 2, 4, 7, 10	6, 13, 15, 11			
An 8	1, 2, 6, 7, 10	4, 13, 15		11	1	4, 7, 13, 11	10, 15	2	6

COMPARISON WITH OTHER SCIENTISTS

Data on the earlier groups are recapitulated in Table 15.[4] The average length of the responses to each card differs considerably, with the social scientists giving significantly longer stories ($p < .05$) than the other two groups, who do not differ materially from each other. This is in accord with the generally greater verbal productiveness of the social scientists and undoubtedly associated with the fact that more of them

---

[4] Some minor discrepancies between the totals shown here and in the tables published in the earlier monographs are accounted for by slight changes in scoring practice; e.g., earlier, if a story seemed to fall equally well into two categories, it was entered for both; now it is entered for one only.

TABLE 15

COMPARISON OF TAT DATA FOR THE THREE GROUPS

	Biologists		Physical Scientists		Social Scientists	
Total number of stories	176		169		200	
Ave. length of stories* per man	12.8	±1.8	13.8	±0.9	20.6	±2.6
	N	%	*N*	%	*N*	%
S or SD	142	81	152	90	185	93
Ds or D	34	19	17	10	15	7
Unusual stories	22	12	25	15	31	15
Unusual twists	37	21	30	18	31	15
Time range complete	23	13	46	28	74	37
Past omitted	129	72	92	55	105	22
Future omitted	69	39	84	50	83	41
Present only	58	33	55	33	62	31
No definite outcome	82	47	79	47	81	41
Success	54	31	54	32	67	34
Defeat	21	12	26	15	31	15
Unresolved tension	19	10	10	6	21	10
Outcome certain	35	37	54	60	70	59
Outcome probable	45	48	34	38	43	36
Outcome possible	14	15	2	2	6	5
Level.						
Concrete-factual	103	58	81	48	96	48
Endopsychic	47	27	62	37	82	41
Symbolic	8	5	8	5	5	2
Past or mythical	4	2	3	2	7	3
Make-believe or dream	6	3	9	5	9	5
Conditional	8	5	6	4	1	1
Tone:						
Indiff., detached, contempl.	13	7	18	11	6	3
Cheerful, serene	16	9	37	22	21	10
Unhappy, tense, anxious	114	65	89	53	127	64
Morbid, aggressive	7	4	14	8	24	12
Melodramatic, sardonic	26	15	9	5	22	11
Personal Relations:						
Formal	77	44	79	46	90	45
Emotional	58	33	42	25	71	36
None, or not indicated	41	23	48	28	39	19
Presses:						
Friendly	44	25	46	27	43	21
Unfriendly	25	14	14	8	46	23
Impersonal	53	30	49	29	53	27
Internal	33	19	21	12	35	17
?	21	12	39	23	23	12

* Number of lines of typescript.

than of the other groups think verbally (19). The greater length of response may also reflect general testwiseness, but this is hard to check. There is practically no difference among the groups in the proportions of unusual stories or of unusual twists to common stories.

The full time range, which is significantly commoner among the social scientists (chi square for 3 × 2 table gives $p < .01$), may also be related to the social science group's willingness to verbalize at greater length. The biologists omit the past more than the others do. Description of the card or a story relating just to the immediate moment was the only response about a third of the time in all groups. The prodding possible in usual clinical practice is clearly inappropriate with these subjects.

There are no differences among the groups with regard to outcome—the proportions are remarkably similar. But there is a marked difference with regard to the certainty with which any outcome is predicted, the biologists being significantly more restricted in this regard than the others.

In all groups the major levels are concrete-factual and endopsychic, with the biologists giving more stories characterized by the former. This accords with other aspects of their general attitudes, such as their greater interest in form on the Rorschach and their generally better emotional control, or emotional flatness. There are no marked differences in tone, in personal relations, or in assignment of presses among the groups.

In these comparisons what is most noticeable is the great over-all similarity. However, analysis of content shows more striking differences. The biologists are the only group whose TAT protocols give any indication of particular meaningfulness to them of the paternal role. Both biologists and physicists are much less interested in interpersonal relations generally, and more inclined to handle them in distance-getting ways than are the social scientists, although many of these are uneasy about them. But the unease is of a different sort and a manifestation of a considerable concern with such relations, rather than a dislike for them. Both biologists and physicists show a considerable independence of parental relations, and without guilt, particularly in the case of the physicists, whereas the social scientists show many dependent attitudes and much rebelliousness, accompanied frequently by guilt feelings. The attitudes of helplessness so noticeable among the social scientists are much less common in the biologists and physicists. The biologists are definitely more restrained than the other two groups in their expression of aggressive attitudes; the social scientists are the freest in this respect.

What is most striking about these results, however, is the fact that the TAT rarely gives any indication that the subject is a man of considerable attainments. Sometimes, some amount of drive is shown, but for the most part this is not very evident in the stories, nor is there any clue in them as to what has made it possible for these men as a group to have achieved as conspicuously as they have.

IX. THE RORSCHACH METHOD OF PERSONALITY DIAGNOSIS

The Rorschach was given and scored according to the directions by Klopfer and Kelley (9), and also by the Munroe Inspection Technique (13). The latter system makes the results easier to handle as a group and makes some allowance for variation in response total.

Knowledge of the Rorschach test varies in the group from none to fair acquaintance. Three have read none of the literature. There are three with some scoring experience—an anthropologist who was administered and scored a great many, a psychologist who has scored a few, and one who had a 10-day course in 1941 but has never used the test. The rest have read varying amounts of the literature—usually just enough to permit them to

TABLE 16

Rorschach Data

Subj.	*R*	*M*	*FM*	*m*	*k*	*K*	*FK*	*F*	*Fc*	*c*	*C′*	*FC*	*CF*	*C*	*W*	*D*	*d*	*DR*	*S*	*P*	*O*	*T*
Ps 1	40	4	4	0	1	0	1	16	3	1	0	6	4	0	10	19	1	7	3	4	4	0
Ps 2	38	3	4	0	0	1	0	19	2	2	0	3	4	0	8	25	4	0	1	9	1	0
Ps 3	80	9	8	9	0	0	1	43	5	2	1	1	1	0	8	38	11	20	3	6	41	0
Ps 4	115	6	15	3	0	1	4	65	6	5	3	4	1	2	7	40	9	32	27	8	57	0
Ps 5	32	2	5	0	2	0	0	10	7	2	0	2	2	0	18	8	2	4	0	5	9	0
Ps 6	92	7	9	9	3	0	0	35	13	3	1	6	6	0	24	39	14	14	1	6	33	1
Ps 7	18	1	0	0	0	0	0	10	1	3	0	2	0	1	6	9	0	3	0	2	1	0
Ps 8	36	2	8	2	0	0	0	15	5	0	2	1	1	0	11	22	2	1	0	6	2	4
Ps 9	53	7	5	1	3	1	0	21	5	1	2	3	4	0	13	28	4	5	2	7	5	0
Ps 10	186	7	16	16	8	1	5	59	24	16	2	13	17	1	49	65	18	46	6	9	72	0
Ps 11	67	16	12	3	2	1	3	11	5	0	3	8	3	0	9	24	8	25	1	5	23	0
Ps 12	76	15	12	1	2	0	2	24	7	1	1	7	4	0	8	35	7	21	5	7	20	0
Ps 13	38	6	4	1	4	0	1	7	4	0	1	4	6	0	17	17	0	2	2	6	6	0
Ps 14	53	5	3	1	1	0	1	26	4	3	0	5	2	2	8	28	7	8	1	4	9	0
An 1	116	7	9	3	4	0	2	67	16	0	2	3	2	1	12	46	19	36	3	6	40	11
An 2	121	18	22	13	3	1	2	33	12	6	0	5	6	0	43	47	10	19	4	6	51	5
An 3	105	6	12	6	4	2	0	40	15	5	1	8	6	1	19	43	18	20	3	6	25	7
An 4	12	2	2	0	0	0	0	4	1	0	0	0	3	0	8	3	1	0	0	3	1	0
An 5	100	5	12	3	4	0	0	30	19	2	8	6	11	0	24	42	6	18	10	7	52	11
An 6	10	4	0	0	0	0	0	1	2	0	1	0	2	0	10	0	0	0	0	3	1	0
An 7	45	11	6	0	0	0	0	13	7	0	0	8	0	0	8	26	5	6	0	6	8	0
An 8	40	5	5	0	0	0	0	13	4	5	0	6	2	0	14	17	1	6	2	7	10	2

include a brief statement, in one of their courses, about the intent of the test.

The usual test data are presented in three tables. Table 16 gives absolute values for all determinants, Table 17 percentages and various other information, Table 18 the content of the responses. The Munroe Inspection Technique date are given in Table 19.

Discussion of the results is organized around the Inspection Technique data in Table 19. Throughout the table minus entries indicate a deficiency, and plus entries an excess of the item (the number of each indicating degree) relative to the total protocol of each individual. Excessively poor or vague forms are noted as *B* or *V*, either with respect to whole or original responses in particular, or to all responses in general. Checks indicate refusal of a card or presence of shading or color shock. Other entries are explained later. Item by item discussion obscures the interdependence of all items in interpretation; I have kept these relations in mind but to have made them

TABLE 17

Additional Rorschach Data

Subject	*R*	*W*%	*D*%	*Dr*%	*F*%	F^1+%	F^2+%	*A*%	Last 3	*T/R*	Ave. *RT*	*RT* Range	*d*%	*S*%	*O*%	Non-*F* Dom. %
Ps 1	40	25	48	18	40	94	88	45	40	18″	8″	5–19″	3	8	10	13
Ps 2	38	20	66	0	50	90	76	42	34	?	3.9	1–10	10	3	3	19
Ps 3	80	10	48	25	54	82	51	31	33	?			14	4	51	10
Ps 4	115	6	35	28	56	83	76	36	34	20″	4.5	1–8	8	23	49	10
Ps 5	32	56	25	13	31	80	72	31	27	64″	26.4	9–78	6	0	28	19
Ps 6	92	26	43	16	38	80	75	33	44	22″	3.4	1–13	15	1	36	15
Ps 7	18	33	50	16	56	100	78	33	28	33″	8.3	2–27	0	0	6	22
Ps 8	36	30	61	3	42	93	95	47	45	30″	11.1	4–22	6	0	47	3
Ps 9	53	24	53	11	40	95	84	35	43	50″	14.9	4–30	8	4	10	15
Ps 10	186	26	35	19	32	83	66	23	38	19″	1	1	10	3	39	28
Ps 11	67	13	36	37	16	100	82	27	45	42″	16.7	7–28	12	2	34	10
Ps 12	76	10	46	28	32	80	71	35	33	35″	6.3	2–13	9	6	26	10
Ps 13	38	45	45	5	18	56	56	23	50	?	8.0	4–17	0	5	16	26
Ps 14	53	15	54	16	49	81	79	34	47	36″	14.2	3–27	14	2	17	13
An 1	116	10	40	32	58	90	90	32	39	44″	11.2	4–23	16	2	35	4
An 2	121	35	39	14	27	79	74	31	33	18″	4.9	1–13	8	3	42	16
An 3	105	18	41	21	39	80	70	28	35	27″	9.4	4–20	17	3	24	19
An 4	12	67	25	0	33	100	75	33	42	25″	5.9	2–10	8	0	8	25
An 5	100	24	42	18	30	67	77	34	44	33″	10.0	2–37	6	10	52	18
An 6	10	100	0	0	10	110	80	10	30	96″	11.0	3–25	0	0	10	20
An 7	45	18	58	13	29	100	97	51	40	29″	13.8	7–23	11	0	18	0
An 8	40	35	43	15	31	100	80	31	43	37″	2.0	1–4	3	5	25	18

TABLE 18

CONTENT IN THE RORSCHACH

Subject	*H*	*Hd*	*A*	*Ad*	*AObj*	*At*	*Aat*	*Sex*	*Sci*	*Obj*	*Pl*	*N*	*Geo*	*Art Des*	*Embl Mask*	*Blood*	*Expl Fire*	*Fount. Cloud Wind*	*X-Ray Photo*	*Clothes*	*Food*	*Mud Ink Rock*	*Ice Snow*	*Arch*	?	Other
Ps 1	4	3	14	4	2	2	1	0	0	3	1	1	1	0	2	1	0	0	0	0	1	0	0	0	0	0
Ps 2	3	2	12	4	3	0	2	1	1	1	1	0	2	0	0	1	0	1	0	1	2	0	0	0	0	eyes 1
Ps 3	6	11	20	5	0	3	0	1	0	22	2	1	3	0	0	0	0	2	0	2	1	1	1	0	0	0
Ps 4	6	7	26	16	3	0	3	1	0	19	0	6	3	16	0	0	1	0	0	0	0	0	0	5	0	star 1
Ps 5	3	1	5	5	2	2	1	0	4	2	2	1	4	0	0	0	0	0	0	0	0	0	0	0	0	0
Ps 6	7	8	16	15	1	6	3	4	0	5	6	2	5	2	2	1	2	2	0	1	0	1	0	0	3	0
Ps 7	1	3	5	1	1	2	0	0	0	0	1	0	0	3	0	0	0	0	0	0	1?	0	0	0	0	0
Ps 8	2	2	14	3	2	7	0	0	0	1	1	0	0	0	3	0	0	0	0	0	0	0	0	0	0	0
Ps 9	7	5	12	6	2	0	1	0	0	3	1	2	5	2	0	0	2	0	0	1	0	2	0	0	0	notes 1, paper 1
Ps 10	7	9	25	18	5	5	5	9	2	29	21	9	11	14	2	1	3	1	0	1	5	0	0	0	0	sym. 1
Ps 11	13	6	10	9	5	0	0	1	1	3	5	3	1	0	1	1	0	1	1	3	2	0	0	0	0	0
Ps 12	18	14	20	4	2	1	0	1	1	6	1	4	0	0	0	0	0	2	0	2	1	0	1	0	0	0
Ps 13	5	0	9	0	2	2	0	1	1	2	2	1	6	3	0	0	1	0	1	0	0	1	0	0	0	0
Ps 14	5	1	12	6	1	5	1	1	1	2	2	2	5	0	1	0	2	0	0	1	2	3	0	0	0	0
An 1	7	28	23	14	2	11	4	4	2	7	3	2	3	2	0	0	0	1	1	0	0	0	2	0	0	0
An 2	19	9	23	14	3	5	0	4	5	13	4	3	11	2	0	0	2	0	0	0	4	0	0	0	0	0
An 3	8	9	20	9	6	13	3	9	0	3	0	3	9	1	1	1	2	2	0	0	1?	2	0	0	0	sym. 3
An 4	3	1	3	1	1	1	0	0	0	0	0	0	0	2	0	0	0	0	0	0	0	0	0	0	0	0
An 5	8	8	19	15	2	8	4	9	0	3	10	0	4	3	3	1	1	0	1	0	0	0	0	0	2	0
An 6	4	0	1	0	3	0	0	0	0	0	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0
An 7	11	4	14	9	3	0	0	0	0	3	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
An 8	5	2	9	4	2	2	1	4	0	2	1	1	1	0	1	0	0	0	0	1	0	4	0	0	0	0

explicit in each instance would have tripled the length of this section.

Total number of responses (*R*) ranges from 10 to 186; the mean for psychologists is 66, for the anthropologists 69. These means are both very high, the usual total being 30 to 40 (3, 9). This group tends also to make more than usually rapid responses, time per response (*T/R*) averaging 33 seconds for psychologists and 39 seconds for anthropologists. Only one subject refused any card.

Location entries. There are 4 entries referring to the portion of the blot in which concepts were seen and the sequence in which these were used. Use of the whole blot is recorded as *W*, of unusual areas as *Dd*, and of white space as *S*. Succession (*Suc*) refers to the orderliness with which different areas are used. In summary there are no major differences between psychologists and anthropologists with respect to their use of locations in the blots, and a particular pattern seems to be characteristic of most of them. They produce an absolutely large, but relatively small, number of whole responses—they can deal with large concepts, can generalize adequately, if sometimes sweepingly, but are usually more interested in smaller, and less often noticed details. They are quite good observers, and tend to look at things which are not likely to strike most people. They are, however, quite casual and unsystematic in the way they go about things, sometimes to the extent of considerable disorganization in the approach. They are so productive, and so many responses occur to them so rapidly that they make no attempt to sort them out, nor do they need to rely upon any technique of procedure to stimulate further responses.

Content. This section refers to the

TABLE 19

Munroe Inspection Technique Record

		Psychologists														Anthropologists							
		Ps 1	Ps 2	Ps 3	Ps 4	Ps 5	Ps 6	Ps 7	Ps 8	Ps 9	Ps 10	Ps 11	Ps 12	Ps 13	Ps 14	An 1	An 2	An 3	An4	An 5	An 6	An 7	An 8
	R	40	38	80	116	33	92	18	36	53	186	67	76	38	53	116	121	105	12	100	10	45	40
	T/R	–			–	+					–						–	–	–		+	–	
	Refusal							√ √															
LOCATION	*W*			–	–	+				–		–	–	+100		–	BV		+		+		
	Dd	+		+	++	+	+	+		+	+	++	++		+	++	+	+		+		+	+
	S	+			++															+++			
	Suc		l	ll	ll	l	ll		l	ll	ll	ll	ll	l	ll	ll	ll	ll		ll		l	
CONTENT	*P*																						
	O			+B	+		+				+	B				+	+B	BV		++			BV
	At						+		+		+				+	++	+	++		++			++
	Range		+	–	t	+	+, t		t	+, t	+, t	+	+, –		+	+, –	+	+t	–, t	+		–	+
FORM	*F%*		+	+	+			+								+					–		
	F		V	BB	BV	BV	BV				BV	B	B	V			VB			BB			
SHADING	*ShSh*	√				√ √	√ √	√ √	√	√		√ √	√					√		√ √	√		
	FK, Fc																						
	c							++			+				+			+					+
	C																			+	+		
	K, k									+	+			+			+	+					
MOVEMENT	*M*	r	rr	db	r			– –	–			++r	++drlil			rd	+B	rlil	r		+	+	
	FM				+	+		–	++		+										–		
	m			+			+				++						+				+		
	Sm							–													++		
COLOR	*C Sh*	√	√ √		√		√ √		√ √	√ √				√ √	√	√ √			√ √	√	√	√	
	FC			–					–										– –		– –		
	CF							–											++		+	–	
	C														+								
	SC			–										+							++		
	C:M		+	– – –				++	U – –		+			+	+								+
	ITS	6	9	18	15	9	13	15	13	10	15	13	13	8	9	14	14	14	11	17	16	7	7

nature of the responses, rather than to the formal aspects, to what they see, rather than how they see it. Entries are for popular responses (*P*), original responses (*O*), anatomy and sex responses (*At*), and range of types of content. (t entries indicate thematic perseverations.)

All of the social scientists give an adequate number of popular responses. A number of them tend to considerable use of original responses, about which they are likely to be rather uncritical. Anatomy and sex responses are relatively common, and are used by almost half of the group to an amount considered excessive. The range of responses is a rather curious and interesting one, since it is cusstomarily very broad in terms of number of categories, and at the same time may show stereotyping or restrictedness of some sort. This is more often in terms of particular individual perseverations of themes, but the restriction may be in terms of excessive use of animal and human responses. The wide range would be associated with their general productivity and must also indicate a pretty general receptivity. That it does not mean an undesirable diffusion of interests would seem to follow from the fact that they are also restricted, as well as from their actual behavior. The frequent emphasis on human responses may well have vocational significance.

The content categories (human, animal, plant, geography, etc.) utilized by the two social science groups are very similar. A comparison of the number in each group using each of 26 different content categories yielded a rank correlation of + .88. The largest rank differences were for the categories food and clothes, these ranking 12th for the psychologists and 19th and 23rd for the anthropologists. Chi square for frequencies is 5.51 for clothes and 3.44 for food; for p of $< .05$ chi square is 3.8.

Form. Form responses are those in which the concept is determined entirely by the outline of some blot area. Check-list entries refer to the relative number of these ($F\%$) and to the accuracy of form for all concepts (not just those scored *F*). Form quality is an indicator of the general soundness of thinking, and more extensively of the soundness of the subject's contact with reality in general. Half of the group do not rate very high in this respect, since they receive entries for excessive use of poor or vague forms. The point may be the nature of the "reality." Psychologists are generally less concerned with what may be called external reality than they are with inner realities (e.g., motivation) and must often disregard the apparent reality and search further. (Is this rationalization?) This may be less true of anthropologists. Whatever the explanation, the fact is that social scientists are relatively unconcerned with formal qualities. This is also shown by the 15.1% average of responses which are not dominated by form. (See Table 16.)

Shading. Shading responses are those in which the tonal quality of the blots is used as texture or vista. Shock is scored when disturbances in the level or time or quality of responses appear on the shaded cards. It is supposed to indicate serious anxiety. I have suggested that it may be a reflection of an insecurity related to early failures, or loss of interpersonal relations which have been accepted or somehow coped with; and that it is not necessarily, by itself, a serious indication (see 23). It occurs in about 50% of this group and is severe in about half of those who show it at all. There are only scattered entries for excess of any particular variety of shading responses and these are not of special importance in the group picture. These groups seem to have fairly effective techniques for handling anxiety.

Movement. Responses are scored for movement when humans (*M*), animals (*FM*), or inanimate objects (*m*) are seen in motion or in attitudes of tension. The *M* responses are supposed to indicate stabilization factors and resourceful capacities, but I think their primary meaning is simply interest in persons (or self) as persons. Excess suggests too much self-preocccupation and the quality of these responses is particularly important. (The letter entries, *B, r,* etc., indicate inadequacies here.) It is striking that only two psychologists and two anthropologists give optimal numbers of good *M* responses.

Some of the anthropologists start a response with free action and then tone the action down so that it becomes very restricted. This is not characteristic of this subgroup generally, and does also happen among the psychologists although less frequently. It suggests as a possible interpretation a need to repress too direct an interest in persons. Anthropology would be a good vocation for those who feel this way, since the interest in persons can be followed in a somewhat depersonalized way. To some extent, this is also true of experimental psychology.

In summary, the use of human movement in these subjects would indicate consistent interest in persons, but an interest which has been frequently restricted in some way and which is sometimes carried to extremes. The subjects tend to excess, rather than deficiency, in any movement category, but not to overall excess in the whole movement area. It is not the picture that they are generally self-absorbed, but rather that they have considerable empathic capacity.

Color. Color entries refer to the way in which the subject makes use of color in developing his concept. Color shock is analogous to shading shock; I have found it of little significance in these groups and recent work throws some doubt on its usefulness. Color responses, generally, refer to the subject's desires for contact with other persons and his method and capacity for implementing these. In general the picture for the social scientists is a well-balanced, one (*FC* is usually larger than *CF* and there are very few *C* responses). The group as a whole shows a fairly rich reactivity to immediate external stimulation, with good emotional control and without impoverishment of reactivity. In individual analysis the content of the color responses is also of importance.

The last entry in the table (*C:M*) refers to the relative numbers of responses using any color or movement. There is much greater tendency in this group to emphasis on color ("plus" entries).

The Inspection Technique Score (ITS). This measure (the total number of entries for any one person with each part of a complex entry counting as 1) is a very rough measure of adjustment. In my experience it correlates adequately with clinical adjustment of the subject rather than with social adjustment. (See reference 23 for fuller discussion.) It shows, that is, what the amount of stress has been rather than the degree of expression of it in behavioral terms. Obviously the lower the score the better the adjustment. Range in this group is 6-18, with an average of 11.4 for the psychologists and 12.5 for the anthropologists. In spite of some high scores it is clear that all of these subjects are functioning adequately, to say the least. For some of them, however, this appears to be at

TABLE 20

RORSCHACH COMPARISON AVERAGES FOR DIFFERENT SCIENTIFIC GROUPS

Entry	Biologists	Physicists	Social Scientists
N	20	19	22
R	22.1	33.7	67.0
W	8.7	9.9	15.2
D	9.6	14.3	28.2
d	1.2	2.3	6.6
dr	2.3	6.3	13.3
S	0.2	0.9	3.4
P	5.2	4.9	5.8
O	4.8	6.7	21.6
F%	43.1	40.9	36.9
A%	42.5	39.6	32.9
F+'%	83.8	85.6	86.9
Non-*F* dom%	7.0	17.4	15.1
T/R	36.9	38.9	30.8
RT	16.7	14.1	8.7
ITS	8.9	11.2	12.1
Age	51.2	44.7	47.7
M	2.6	2.9	6.7
FM	3.1	3.6	7.9
m	0.45	1.3	3.2
k	0.55	0.8	1.9
K	0.10	0.5	0.4
FK	0.10	0.5	1.0
F	9.7	14.3	25.5
Fc	2.3	2.9	7.6
c	0.15	0.8	2.6
C'	0.15	0.5	1.3
FC	2.2	2.4	4.6
CF	0.75	2.9	3.9
C	0	0.2	0.3
Sum C	1.8	4.3	6.7

quite a cost and it is this that these scores indicate.

Qualitative aspects. Perhaps the comments most frequently found in the individual analyses refer to the general productiveness of the men in these groups, to their rather uncritical attitudes, and a sort of haphazard use of rational controls—that is, that they can be rational when they wish to be but generally feel no compulsion to make a point of being so. A very great sensitivity is also extremely noticeable in almost all of the records and it usually implies a great awareness of other persons. It may sometimes result in an easy irritability, but I think more often not. In most of the protocols, there is evident fairly free aggression, which is clearer and stronger generally among the anthropologists, and more obviously oral among the psychologists. There are a number in which there seems to be a strong consciousness of hidden things, but this is not always accompanied by anxiety. Most of the subjects are fairly warm persons, but this is not always the case. Conflict over dominance and authority is common. There is also evidence, particularly among the psychologists, of needing to hold and to feel nurturant attitudes.

COMPARISON WITH OTHER SCIENTISTS

The Rorschachs from these 22 social scientists can now be compared with those obtained from the 20 biologists and the 19 physical scientists previously studied. Means for each group for the various Rorschach determinants and some other data are given in Table 20.

The great difference in average number of responses between the social scientists and the others makes direct comparison for mean frequencies of little value, and most percentages are also so affected by total number of responses that they can also not be fruitfully compared. The $F + \%$ (the percentage of responses which are good form) and the non-*F* dominated responses (total of *mF*, *cF*, *CF*, *C*, etc.) as well as the ITS are not so affected. Analysis of variance of these scores gives the following results:

	F	p
R	13.33	$<.001$
$F+\%$	.52	.10
non-F dominated %	5.84	$<.01$
ITS	9.58	$<.001$

Referring to Table 19, it appears that the social scientists are significantly more productive on the Rorschach; that the biologists use relatively fewer responses not dominated by form than the others; and that the biologists, by Munroe's measure, are definitely the best adjusted. The two latter differences are certainly related, since the nature of the adjustment shown by the biologists is one in which rational control and caution are emphasized.

In Table 21 are listed the checklist entries which show some differentiation among these groups. Only entries for which chi square has a p of nearly .05 or less are given. Although the checklist scoring is adjusted for length of the individual protocol, some of the observed significant differences are related to difference in length. Succession cannot be scored in short protocols, and range is more likely to be great in longer protocols. Where very few M are given no entry for restricted M (r) can be made. The contrasts between physicists and social scientists in use of W, between the social scientists and the others in use of M, and between the physicists and biologists in use of CF do seem to be sustained by this analysis.

Differences between the scientist groups in content categories can be expressed over-all by the use of rank correlations (on 24 categories). These are: biologists-physicists + .739; biologists-social scientists + .713; and physicists-social scientists + .769. These are not high. Major differences are greater use by biologists of the categories science, animal anatomy, and abstract; by physicists of art and design and emblem; and by social scientists of clothes and food.

Some qualitative differences may be indicated also. The biologists are the least freely aggressive; the social scientists, particularly the anthropologists, the most so, and with greater likelihood of oral elements. There are great differences between the biologists and physicists in their handling of anxiety, but the social scientists show no consistency in this respect.

In the over-all picture the similarities are greater than the differences. This is

TABLE 21

COMPARISON OF CHECKLIST ENTRIES FOR DIFFERENT GROUPS OF SCIENTISTS

(3×2 tables, except as noted.)

Entry	Frequency for—				
	Biologists	Physicists	Social Scientists	χ^2	p
N	20	19	22		
$W+$	3	8	4	12.02*	$<.01$
$W-$	0	2	6		
Suc, l or ll	6	7	17	11.05	$<.01$
$O+$	2	1	7	6.13	$<.05>.02$
Range+	1	2	13	19.36	$<.01$
F, BV	3	7	11	5.76	$>.05$
Mr	6	2	10	6.00	.05
$M+$	1	0	5	17.37*	$<.01$
$M-$	10	9	2		
$CF+$	0	7	2	14.28*	.01
$CF-$	7	1	2		

* W, M, and CF are checked by 3×3 tables, the rows being +, no entry, and −.

TABLE 22

COMPARISON OF INDIVIDUAL AND GROUP RORSCHACHS OF SOCIAL SCIENTISTS

	Means		t	p
	Individual	Group		
N	22	129		
Age	47.7 ± 1.38	41.1 ± 0.87	4.04	$<.01$
R	67.0 ± 9.08	42.0 ± 1.65	2.66	$<.01$
ITS	12.1 ± 0.74	8.2 ± 0.35	4.73	$<.01$
$F+\%$	86.9 ± 2.56	83.4 ± 1.71	1.13	$>.05$
Non-F dom.	15.1 ± 1.53	10.1 ± 0.58	3.05	$<.01$

Checklist Entries

Entry	Frequencies for—		χ^2	p
	Individuals	Groups		
$W-$	6	10	7.56	$<.01$
$Dd\%$	17	60	7.12	$<.01$
$O+$	7	0	35.76	$<.01$
O, BV	5	4	12.73	$<.01$
At, Sex	9	15	12.05	$<.01$
Range+	13	23	17.53	$<.01$
Range− or t	11	32	5.86	$.02>p>.01$
F BV	11	37	3.94	$.05>p>.02$
$C:M+$	6	12	7.65	$.05>p>.02$
$C:M-$	2	27		

to be expected from the fact that there is considerable heterogeneity within the separate groups, and from the fact that these men are all functioning adequately.

COMPARISON WITH GROUP RORSCHACH STUDIES

The group Rorschach was given to 104 psychologists and 25 anthropologists, members of university faculties. A full report is given elsewhere (23) but the data for the total group are used here for comparison with the men studied individually. There is a major difference in the proportions of types of psychologists in the two groups. Among those taking the group Rorschachs there were 25 experimentalists, 33 clinicians, 27 working in social, child, and personality psychology and 19 in industrial and statistical. Some differences were found for these subgroups, but these are relatively few.

Table 22 gives comparative data from the individual and group studies. Age difference is a result of the selection of the men for individual study. The very large and significant difference in mean number of responses makes comparison, except by the checklist, possible for only a few items. Differences in ITS and in non-F dominated % are significant.

Only 9 checklist items show significant differences; these, if they can be accepted at face value, would indicate that the eminent group, in addition to its greater productivity, used fewer whole and more unusual detail responses, were both more original and less controlled, produced a wider range of responses, including more anatomy and sex responses and more concept-dominated series of responses, and finally, tended to proportionately more color than movement responses. These can be subsumed under a general attitude of greater productivity and reactivity, more originality, and less control.

There is quite close correspondence in use of content categories, rho being + .88. There are no major differences among responses most commonly given, except the anatomy responses.

Comparison of eminent biologists with other biologists produced a different picture—the more eminent men had rather better controlled, if somewhat more restricted, protocols. Here the opposite is true. In the case of physicists no important differences were found.

Comparisons on the checklist between all of the eminent men and all of those who took the group Rorschach show a few major differences which are not related to the longer individual protocols. Among these the most marked are the greater use of unusual blot areas by the eminent men, and their freer use of anatomy and sex responses and of perseverating responses.

limitation

X. DISCUSSION

The direct study of eminent men raises numerous and very difficult problems. One clearly does not have the complete freedom of a biographer writing centuries after the lifetime of the subject. But these difficulties are more than compensated for by the value of direct clinical and, test data. In the first study of such a nature, much time must be spent in exploratory work and the first monograph pointed out that at this stage, "All that one can hope for in such work is to get some idea of the nature of the relationships, the points at which a direct attack can be made, and the sort of tools to use" (21, p. 1). I feel that this has been accomplished. Before explicit discussion is presented, however, something should be said about the limitations of the study.

In the first place the sample is small in absolute numbers although relatively very large. The subjects are the best research men in each field and they comprise a high percentage of the men who could be so designated. The conclusions drawn, however, apply directly only to the first-rate scientist, and only indirectly, and with some qualifications, to scientists generally. The group Rorschachs have offered useful confirmation, however.

We lack comparable groups in nonscientific vocations. A more serious limitation is the lack of any control group of relatively unsuccessful scientists, men who had the training and appeared to have the promise, but who have produced little or not at all in research. This is the next most important step and a prerequisite to the satisfactory development of hypotheses about choice of science as a vocation and success in it. One cannot always be certain whether the situations noted in this study refer to choice of vocation or to success, or to what extent they are affected by high frequency of a middle-class socioeconomic background.

As in all research with people, the complexity of the situations encountered makes the determination of direct causal relations practically impossible. What has been accomplished, however, is not only the accumulation of test data on a hitherto practically unstudied group, but also the identification of situations which recur with high frequency.

It is evident that the family backgrounds of the 64 scientists studied are by no means randomly selected with respect to the population at large. According to census reports for 1910, only 3% of the gainfully employed men in the country were professional men. In this group, however, 53% of the fathers of the subjects were professional men. One-eighth of the group came from farm homes; and the fathers of 31% were in business, many of them owning their own. Only two fathers were skilled laborers. None of the scientists came from homes in which the father was an unskilled laborer and none came from families of very great wealth. Cattell and Brimhall (5) in 1921 found a 51% incidence of professional fathers for the 66 leading scientists they studied.

What seems to be the operative factor here is that in practically all of these homes, whatever the occupation of the father, learning was valued for its own sake. Its concomitants in terms of possible better income or social position were not scorned, but it was rare for these to be the most important. This certainly was a major factor in the facilitation of intellectualization of interests. In my opinion this, rather than the probable associated intellectual levels, is the important aspect here. "Overintellectualization" may be a middle-class characteristic and it may interfere with libidinal development in other spheres, as some psychoanalytic writers have pointed out. Yet it seems to me doubtful whether one can develop the sort of intense personal involvement which is characteristic of these scientists without some degree of this, if a channeling of energy in one direction means a lessening of it in others. There is a serious problem here. Unquestionably overintellectualization is frequently a technique for escaping emotional problems, especially those bound up in interpersonal relationships, but it is not necessarily so. I believe it is possible to concentrate upon intellectual activities without having a relatively sterile life emotionally, but we certainly have not developed educational techniques which foster this.

I have reported a greater than chance incidence of first-born among these eminent men. The problem of birth order is an extremely tricky matter statistically, and I would not be inclined to pay much attention to this finding in a group of 64 were it not that Cattell and Brimhall reported the same finding in a group of 855 scientists. It could be argued that the point here also is that intelligence levels are higher in the first-born, for which there is some evidence, but it seems much more probable that both of these facts are results of the same cause, whatever that is.[5]

Certain aspects of the data offer evidence on the basic importance of the need to achieve, or to keep independence, which is so well met by a career in research. There are no Catholics in the group. The Protestant churches to which all but five of the scientists' families belonged have varying degrees of insistence on the authority of the church over its members' interpretations of life, but all but three of these subjects have dismissed organized religion as a guide and usually had done so by late adolescence. In this respect, also, they have achieved independence. The dearth of Catholics in research science is corroborated in other studies (11, 29) and the Wesleyan survey found that production of scientists from intellectual channels were sought to alleviate it must be in large part because of the family background, but there is no question that the research aspect is of more importance than the general intellectualization.

---

[5] Two hypotheses come to mind. One is that first-born are likely to be overprotected, especially in families where social status is important. These men, then, may be compensating in terms of seeking greater independence. The other hypothesis is that eldest sons may have more responsibility for themselves and have it earlier than is the case for other children; they also have been spared the discouragement of not being able to compete with those just older. Hence they are just continuing an early pattern of independence.

The question also arises as to why one subject chose one field of science and others chose other fields. Apart from the often overlooked matter of necessary contact with the field, there is some further evidence from the study. The problem of coping with early affectional loss has been mentioned. It would also appear that there are some, particularly among the experimental physicists, who seem early to have formed direct relationships with objects rather than people, not compensatorily. In others, a generalized anxiety, of unknown cause, and possibly only an exacerbation of normal anxiety, is alleviated by concentration on a particular field. For example, I know biochemists who seem to me to live in a very dangerous world—they are always conscious of the presence about them of dangerous micro-organisms. They tolerate this in part because they are able to manipulate these organisms to some extent professionally. I am sure, however, that to them psychologists live in an equally dangerous world, surrounded by irrational emotional people, a situation which they would find quite intolerable.

The social scientists stand apart as having been more concerned at an earlier age, about personal relations (or as being willing to tolerate this concern as such, without translation). This may reflect an unconscious uncertainty over the consciously felt superiority that characterized half of the psychologists and most of the anthropologists. It is also certainly related to their difficulties in freeing themselves from their parents. The other groups seem to have been able, fairly early, to work out an adaptation not nearly so dependent upon personal relations, but rather strikingly independent of them. Certainly psychology to some extent, particularly social psycholgy, and anthropology to a large extent, particularly cultural anthropology, offer an ideal vocation to the person whose conviction of personal superiority is not accompanied by asocial characteristics; they permit a somewhat Jovian survey of their own society as well as others, and maintain the social scientist in a state of superiority just because he is able to make the survey. (This accounts nicely for the observation that some rather paranoid indications in the test material are not accompanied by forms of paranoid behavior, except perhaps as regards their own colleagues.) The experimental psychologists are generally less concerned with people as people, although this is by no means true of all of them. The further observations that a conflict over dominance and authority is common in the group, and that in a number of their homes the mother was dominant indicate the possibility of difficulties in achieving masculine identification.

In this respect it would seem very probable that the physicists, particularly the experimentalists, were able to identify more easily with their fathers than the other groups and hence to follow comfortably a science which has rather more of a "masculine" tinge in our culture than the others do.

It must be pointed out that it is likely that the kind of person who has gone into social science may have had a biasing effect on the theories produced by social scientists, particularly with regard to the desirable or the mature personality. Practically all current psychological theory of development stresses strongly the central importance in any life of the richness of personal relations as a basis for "adjustment." But the data of this study demonstrate, and it seems to me quite conclusively, that a more than adequate personal and social adjustment in the larger sense of an adjustment which permits a socially extremely useful life and one which is personally deeply satisfying, is not only possible, but probably quite common, with little of the sort of personal relations which psychologists consider essential. Many of the biological and physical scientists are very little concerned with personal relations, and this is not only entirely satisfactory to them, but it cannot be shown always to be a compensatory mechanism (nor are compensatory mechanisms necessarily undesirable).[7] It can also apparently be satisfactory to others who are closely associated with them. That divorces are so much commoner among the social scientists is of interest in this connection. Problems with masculinity and dominance must be important here; but also, where much more attention and emotion are invested, demands are certain to be greater and more specific, and hence failure commoner.

Another finding of considerable importance is the differences of imagery which are associated with the different fields of science, and which accord with and perhaps explain some of the test data. Briefly, the biologists and experimental physicists tend strongly to dependence upon visual imagery in their thinking; the theoretical physicists and the social scientists, to dependence upon verbalization or similar symbolization in theirs. Nothing is known about the development of these modes of thinking, but it seems probable that they were developed early (they are associated with father's occupation) and played a part in the choice of a science. Further, it was shown that those scientists whose preferred mode of thinking differed from that characteristic for their science also differed in some aspects of their early history, and in the things they did or the ways they went about their work. (This is good reason for not using such a factor selectively—their contributions have a special place.) The domination of the formal qualities of the blots in the biologists' Rorschachs, which the others do not show, is in accord with this, as is the generally much more fluid verbalization of the social scientists.

Doubtless, also, some intellectual factors enter. So far as the test used is a measure of these, it is clear that the theoretical physicists surpass all other groups on both verbal and spatial tests. The experimental physicists are high on the spatial and relatively very low on the verbal test. Psychologists are at about the mean for this total group on all three. Anthropologists are high on the verbal and lowest on both spatial and mathematical. These patternings are probably of importance in selection of vocation—particularly the relatively low nonverbal abilities of the anthropologists and the relatively low verbal ability of the experimental physicists.

I suspect that the verbalization so characteristic of the social scientist has also exerted some bias on his activities. This is probably most obvious in the field of testing where the emphasis still remains on verbal tests, although other tests have come into general use. But psychologists, and educators who are probably much like them in this respect, are in a position which makes possible the operation of this bias to keep out of college many adolescents who are verbally inept but have other capacities of equal value to society, and for whom college could be important. This bias may have affected the development of techniques of teaching and of therapy. The effect in the first is obvious. In therapy it may well be a factor in the common insistence on verbalization of insights as essential to therapy.

---

[7] The fact that a satisfactory life has been achieved on this basis does not mean that a more satisfying one could not be achieved. It is, indeed, a great pity that so many men have less rich lives than they might, but again, I speak from the standpoint of a psychologist. It is, however, probable that a concentration which is basically neurotic will also limit the possible breadth of vocational activity and it will frequently interfere, in the long run, with the man's enjoyment of his vocation, as is witnessed by the not infrequent depressions experienced in the face of the greatest recognition.

That verbalization and intense interest in persons are related has long been noted peripherally. This relation is accompanied by some cultural sex differentiation. Girls test higher verbally than boys; the M-F (or masculinity-femininity indexes) for certain occupations which have culturally a strongly feminine tinge are very different from those with a culturally strongly masculine tinge, and these are also associated with verbalization. See, for example, Table 29 in Strong (27). Quite likely the development of verbalization is higher among those interested in persons, because it is the chief means of communication.

The range of test intelligence in this group is also of importance. All of the evidence confirms Cox's remark: ". . . high, but not the highest intelligence, combined with the greatest degree of persistence, will achieve greater eminence than the highest degree of intelligence with somewhat less persistence" (6, p. 187). Portenier noted that "It would seem then that while there is a positive correlation between psychological test ratings and honor awards, the honor recipients are not limited to students with high psychological tests scores, and many students who make high test scores fail to win honors" (14, p. 499). Clearly a certain degree of intelligence is a necessary condition for a career in research science, but it is not a sufficient one.

The strength of the achievement drive which these men have shown is rarely reflected in the TAT in any direct way, and there are a number of Rorschachs which give no indication that the subject is capable of great accomplishment. Indeed there are a number of subjects for whom none of the test material would give the slightest clue that the subject was a scientist of renown.

There are Rorschach protocols which would occasion no surprise in a clinic for the maladjusted. It is certainly true that those who work only with persons whose lives show considerable disruption seem to have no idea of the extraordinary range of tolerance of difficulty which "normals" show. A number of these men are particularly good examples. It should also be pointed out that for many of these subjects, the career itself has served as a technique for handling the personal problems. In some instances the basic problem has been, in a sense, extrapolated into a more general one, and the subject has then settled down to working on the general problem. This is a very neat and effective method. In other instances, absorption in the career has made possible the encapsulation of the difficulty in such a manner that it can be almost ignored by the subject. The price he may pay for this is another matter. There is nothing in these data to suggest that any measure from these or other projective techniques, or from intelligence tests, would be nearly so adequate in predicting their success as the fact that they worked long hours in graduate school, many more than the course requirements, and that they preferred to work on their own. (But I do not know how many less successful scientists have worked hard and preferred independence.)

Nevertheless the tests have contributed materially to our understanding of what sort of men these scientists are, and have also offered essential clues as to how and why they have become what they are. These now can be followed up in more direct fashion.

It would seem that nearly all educational systems tend to stultify any attempt to learn how to do things or to learn things for oneself. It is much easier to teach dogmas, of whatever variety, to require only rote learning (instead of only so much as is actually necessary as a tool), and to forget that even our tools are only conventions. It is easy to penalize independent thinking. (How much have our "rapid check" tests, helped in this?) It is easier to give assignments in terms of so many pages to be read than in terms of problems to be solved by whatever means can be found. The point is crucial, and it is as important in the elementary and high schools as in college.

Most of these subjects were fortunate enough somewhere along the line to have found a teacher who induced them to find things out for themselves, or who let them do so, or who insisted that they do so because he did not want to be bothered. Once intellectual independence was really tasted, nothing else mattered much pedagogically; bad teaching then was only an irritation. But how many are there who have never learned to rely upon themselves, to find how valid their own thinking may be? Certainty of his own worth is any man's greatest need. Though some of them may find it only there, scientists do find this certainty in science.

XI. SUMMARY

This monograph has presented the life history and test data of 14 eminent psychologists and 8 eminent anthropologists, and compared them with the biologists and physicists previously studied. This summary will omit the comparative material.

Selection was by peer ratings of men presently doing research. Average age is 46.7 for the psychologists, 49.4 for the anthropologists. The majority of both groups came from lower to upper middle-class backgrounds. The economic level was generally higher for the anthropologists. The fathers of half of the psychologists and of three of the anthropologists were professional men.

All of the subjects are married and most of them have children. Average age at marriage was 26.5 for psychologists and 26.1 for the anthropologists. Five of the psychologists and four of the anthropologists have been divorced at least once.

They received their B.A.'s at an average age of 21.4 for psychologists, 22.1 for anthropologists; their Ph.D.'s at an average age of 25.8 for psychologists, 28.6 for anthropologists.

Early interest in literature and the classics was common among both groups, and there were a few with early natural history interests. The psychologists were relatively late in determining upon a profession, largely because psychology was not taught in high school or early college.

Among both groups, particularly the

is probably most obvious in the field of testing where the emphasis still remains on verbal tests, although other tests have come into general use. But psychologists, and educators who are probably much like them in this respect, are in a position which makes possible the operation of this bias to keep out of college many adolescents who are verbally inept but have other capacities of equal value to society, and for whom college could be important. This bias may have affected the development of techniques of teaching and of therapy. The effect in the first is obvious. In therapy it may well be a factor in the common insistence on verbalization of insights as essential to therapy.

That verbalization and intense interest in persons are related has long been noted peripherally. This relation is accompanied by some cultural sex differentiation. Girls test higher verbally than boys; the M-F (or masculinity-femininity indexes) for certain occupations which have culturally a strongly feminine tinge are very different from those with a culturally strongly masculine tinge, and these are also associated with verbalization. See, for example, Table 29 in Strong (27). Quite likely the development of verbalization is higher among those interested in persons, because it is the chief means of communication.

The range of test intelligence in this group is also of importance. All of the evidence confirms Cox's remark: ". . . high, but not the highest intelligence, combined with the greatest degree of persistence, will achieve greater eminence than the highest degree of intelligence with somewhat less persistence" (6, p. 187). Portenier noted that "It would seem then that while there is a positive correlation between psychological test ratings and honor awards, the honor recipients are not limited to students with high psychological tests scores, and many students who make high test scores fail to win honors" (14, p. 499). Clearly a certain degree of intelligence is a necessary condition for a career in research science, but it is not a sufficient one.

The strength of the achievement drive which these men have shown is rarely reflected in the TAT in any direct way, and there are a number of Rorschachs which give no indication that the subject is capable of great accomplishment. Indeed there are a number of subjects for whom none of the test material would give the slightest clue that the subject was a scientist of renown.

There are Rorschach protocols which would occasion no surprise in a clinic for the maladjusted. It is certainly true that those who work only with persons whose lives show considerable disruption seem to have no idea of the extraordinary range of tolerance of difficulty which "normals" show. A number of these men are particularly good examples. It should also be pointed out that for many of these subjects, the career itself has served as a technique for handling the personal problems. In some instances the basic problem has been, in a sense, extrapolated into a more general one, and the subject has then settled down to working on the general problem. This is a very neat and effective method. In other instances, absorption in the career has made possible the encapsulation of the difficulty in such a manner that it can be almost ignored by the subject. The price he may pay for this is another matter. There is nothing in these data to suggest that any measure from these or other projective techniques, or from intelligence tests, would be nearly so adequate in predicting their success as the fact that they worked long hours in graduate school, many more than the course requirements, and that they preferred to work on their own. (But I do not know how many less successful scientists have worked hard and preferred independence.)

Nevertheless the tests have contributed materially to our understanding of what sort of men these scientists are, and have also offered essential clues as to how and why they have become what they are. These now can be followed up in more direct fashion.

It would seem that nearly all educational systems tend to stultify any attempt to learn how to do things or to learn things for oneself. It is much easier to teach dogmas, of whatever variety, to require only rote learning (instead of only so much as is actually necessary as a tool), and to forget that even our tools are only conventions. It is easy to penalize independent thinking. (How much have our "rapid check" tests, helped in this?) It is easier to give assignments in terms of so many pages to be read than in terms of problems to be solved by whatever means can be found. The point is crucial, and it is as important in the elementary and high schools as in college.

Most of these subjects were fortunate enough somewhere along the line to have found a teacher who induced them to find things out for themselves, or who let them do so, or who insisted that they do so because he did not want to be bothered. Once intellectual independence was really tasted, nothing else mattered much pedagogically; bad teaching then was only an irritation. But how many are there who have never learned to rely upon themselves, to find how valid their own thinking may be? Certainty of his own worth is any man's greatest need. Though some of them may find it only there, scientists do find this certainty in science.

XI. SUMMARY

This monograph has presented the life history and test data of 14 eminent psychologists and 8 eminent anthropologists, and compared them with the biologists and physicists previously studied. This summary will omit the comparative material.

Selection was by peer ratings of men presently doing research. Average age is 46.7 for the psychologists, 49.4 for the anthropologists. The majority of both groups came from lower to upper middle-class backgrounds. The economic level was generally higher for the anthropologists. The fathers of half of the psychologists and of three of the anthropologists were professional men.

All of the subjects are married and most of them have children. Average age at marriage was 26.5 for psychologists and 26.1 for the anthropologists. Five of the psychologists and four of the anthropologists have been divorced at least once.

They received their B.A.'s at an average age of 21.4 for psychologists, 22.1 for anthropologists; their Ph.D.'s at an average age of 25.8 for psychologists, 28.6 for anthropologists.

Early interest in literature and the classics was common among both groups, and there were a few with early natural history interests. The psychologists were relatively late in determining upon a profession, largely because psychology was not taught in high school or early college.

Among both groups, particularly the

anthropologists, early feelings of personal or family superiority on a social or intellectual basis were common. Patterns involving overprotection and firm, if not overt, control were frequent, and strong rebelliousness was usual. A number of the subjects still show resentment over family discipline or interference.

All but two of the men came from Protestant homes, none from Catholic homes, and most had some religious training. Only two are now interested in church.

Average raw scores on the verbal test were 57.7 for psychologists, 61.1 for anthropologists. On the spatial test they were 11.3 and 8.2 respectively, and on the mathematical 15.6 and 9.2.

On the Thematic Apperception Test both groups gave relatively long stories, and manifested generally a similar picture. A common theme is of general helplessness in the face of severe problems. There is considerable dependence on parent figures, and a number of stories of unhappiness and guilt with regard to this relation. The group is strongly concerned with interpersonal relations, fairly free in discussing heterosexual ones, and not particularly conventional.

On the Roschach the social scientists are remarkably productive, rather uncritical, and somewhat haphazard in their use of rational controls. They are very sensitive, intensively concerned with persons, rather freely aggressive, and often troubled with conflicts over dominance and authority.

The group Rorschach records of 129 other psychologists and anthropologists have a much lower average number of responses and a significantly better adjustment score. The eminent group used less *W* and more *Dr*, were both more original and less controlled, produced a wider range of responses, including more anatomy and sex responses, and more concept-dominated series of responses, and tended to proportionately more color than movement.

REFERENCES

1. Angyal, A. *Foundations for a science of personality.* New York: The Commonwealth Fund, 1941.
2. Baas, M. L. Kuder Interest patterns of psychologists. *J. appl. Psychol.,* 1950, **34,** 115-117.
3. Beck, S. J., Rabin, A. I., Thiesen, W. G., Molish, H., & Thetford, W. N. The normal personality as projected in the Rorschach test. *J. Psychol.,* 1950, **30,** 241-298.
4. Boring, E. G. Great men and scientific progress. *Proc. Amer. phil. Soc.,* 1950, **94,** 339-351.
5. Cattell, J. M., & Brimhall, D. R. *American men of science.* (3rd Ed.) Garrison, N.Y.: Science Press, 1921.
6. Cox, C. S. *Genetic studies of genius.* Vol. II. *Early mental traits of three hundred geniuses.* Stanford: Stanford Univer. Press, 1926.
7. Guilford, J. P. Creativity. *Amer. Psychologist,* 1950, **5,** 444-454.
8. Henry, W. The business executive: the psychodynamics of a social role. *Amer. J. Sociol.,* 1949, **54,** 287-291.
9. Klopfer, B., & Kelley, D. M. *The Rorschach technique.* Yonkers, New York: World Book Co., 1942.
10. Knapp, R. H., & Goodrich, H. B. The origins of American scientists. *Science,* 1951, **113,** 543-545.
11. Lehman, H. C., & Witty, P. A. Scientific eminence and church membership. *Sci. Mon.,* 1931, **33,** 544-549.
12. Morgan, C. D., & Murray, H. A. A method for investigating phantasies: the Thematic Apperception Test. *Arch. Neurol. Psychiat., Chicago,* 1935, **34,** 298-306.
13. Munroe, Ruth. Prediction of adjustment and academic performance of college students. *Appl. Psychol. Monogr.* 1945, No. 7.
14. Portenier, L. G. Mental ability ratings of honor students. *J. educ. Psychol.,* 1950, **41,** 493-499.
15. Rapaport, D. *Diagnostic psychological testing.* Chicago: Year Book Publishers, 1945.

16. ROE, ANNE. Analysis of group Rorschachs of biologists. *J. proj. Tech.*, 1949, **13**, 25-43.
17. ROE, ANNE. Psychological examinations of eminent biologists. *J. consult. Psychol.*, 1949, **13**, 225-246.
18. ROE, ANNE. Analysis of group Rorschachs of physical scientists. *J. proj. Tech.*, 1950, **14**, 385-398.
19. ROE, ANNE. A study of imagery in research scientists. *J. Pers.*, 1951, **19**, 459-470.
20. ROE, ANNE. Psychological tests of research scientists. *J. consult. Psychol.*, 1951, **15**, 492-495.
21. ROE, ANNE. A psychological study of eminent biologists. *Psychol. Monogr.*, 1951, **64**, No. 14 (Whole No. 331).
22. ROE, ANNE. A psychological study of eminent physical scientists. *Genet. Psychol. Monogr.*, 1951, **43**, 121-239.
23. ROE, ANNE. Analysis of group Rorschachs of psychologists and anthropologists. *J. proj. Tech.*, 1952, **16**, 212-224.
24. ROE, ANNE. Group Rorschachs of university faculties. *J. consult. Psychol.* 1952, **16**, 18-22.
25. ROGERS, C. R. Client-centered therapy. Boston: Houghton Mifflin, 1951.
26. TERMAN, L. M., *et al. Genetic studies of genius.* Vols. I through IV. Stanford: Stanford Univer. Press, 1925-1947.
27. STRONG, E. K. *Vocational interests of men and women.* Stanford: Stanford Univer. Press, 1948.
28. SURVEY RESEARCH CENTER, UNIVERSITY OF MICHIGAN. Productivity, supervision and employee morale. *Human Relations Series,* 1948, **1**, Report 1.
29. VISHER, S. S. Starring in *American men of science. Science,* 1947, **106**, 359-361.
30. WYATT, F. The scoring and analysis of the Thematic Apperception Test. *J. Psychol.*, 1947, **24**, 319-330.

(Accepted for early publication March 2, 1953)

Vol. 67, No. 3 Whole No. 353, 1953

Psychological Monographs: General and Applied

The Development of a Test for the Measurement of Anxiety: A Study of Its Reliability and Validity[1, 2]

M. J. Freeman
North Hollywood, California

A review of the literature on anxiety reveals agreement on the fundamental assumption that fear, frustration, conflict, threat, and apprehension are basic facets of anxiety. Although in many instances a divergence of opinion exists as to the relative importance and primacy of these emotional forces, there is no dispute as to the centrality of anxiety in relation to the development of neurosis.

A. Definition of Manifest Anxiety ("MA")

In this study the term *manifest anxiety* embodies the behavioral responses of fear and apprehension which psychiatrists have observed to be manifest to an abnormal degree in persons suffering from neurosis.[3] It was assumed that these observable manifestations are in effect cues to a nuclear syndrome of anxiety.

This study assumes that anxiety as a pathological manifestation is to be differentiated from normal anxiety which occurs in response to an actual external threat. The following psychiatric requirements for the diagnosis of manifest anxiety were laid down in the present study:

a. Minimum observation period of 30 days by attending physicians, most of whom were psychiatrists.

b. In each instance, medical findings which reported no organic pathology to account for the symptoms.

c. In each instance, concurrence by the head hospital psychiatrist in the diagnosis of manifest anxiety.

B. Specific Purposes of the Present Investigation

If the assumptions in the above sec-

---

[1] This test is now combined with the psychosomatic test which was published separately (3). Since the latter instrument was standardized upon the same subjects as in the present investigation, the final instrument provides a suitable measure of both anxiety neurosis and psychosomatic syndromes.

[2] In view of the fact that this final measuring instrument is based upon an original doctoral research performed at the Claremont Graduate School, as well as upon earlier graduate research at the University of Southern California, the author desires to express grateful acknowledgment to Dr. J. P. Guilford of the University of Southern California and to Dr. Florence Mateer, Dr. F. Theodore Perkins, and Dr. J. E. Caster of Claremont Graduate School, for their respective guidance in the various phases of this research. Thanks are also extended to Dr. Eugene Ziskind, Dr. Harry Roback, and Dr. Karl O. Von Hagen for their excellent cooperation in providing medical facilities for validation of the criterion cases.

[3] The term "manifest" has been selected in order to identify that form of anxiety which is observable, and therefore capable of being identified, by the attending psychiatrist. It is to be distinguished from "latent" anxiety, which would provide a difficult basis upon which to make an objective evaluation. The term "manifest anxiety" therefore involves an overt expression of the inner (nuclear) feelings of abnormal fears and apprehension.

tion are correct, it follows that manifest anxiety, as defined, will not appear in a normal population, and that manifest anxiety will be present in the various forms of anxiety neuroses.

The problem here is twofold: (*a*) to develop a valid measuring instrument of manifest anxiety ("MA"), and (*b*) to differentiate "MA" cases from the normal population.

Specific questions for investigation were as follows:

1. Does the factor of hospitalization have an effect upon the development of manifest anxiety ("MA")?

2. Does sex have any relationship to "MA"?

3. What type of item would prove most successful in differentiating the "MA" criterion cases from the normals? What implications may be drawn from the content of the valid items?

4. Is the psychiatric classification of manifest anxiety a clinical entity capable of objective measurement?

5. Is manifest anxiety nuclear to all forms of psychoneurosis?

C. Development of Test Materials

The following factors have been given careful consideration in order to overcome the common weaknesses of a paper-and-pencil personality test:

a. The elimination of questions or declarative items the answers to which an uncooperative subject could readily falsify; e.g., "Are you nervous?" and "I daydream very little."

b. The construction of items with meaning-tones sufficiently disguised that the subject will be unaware of the diagnostic intent of the items. (This purpose, it is realized, cannot be fully accomplished in every instance.)

c. The formulation of items of a projective character. The test is to bring about projections through the respondent's use of the mechanism of identification. Examples:

Usually when there is pain in one part of the body, other parts of the body also become affected with pain. ( ) Yes ( ) No

The person who presents the more undesirable behavior is
- (1) the one who criticizes people
- (2) the person who lies occasionally ( ) 1 ( ) 2

A person who prefers to be alone is usually
- (1) an unfriendly person
- (2) one who minds his own business
- (3) a home lover ( ) 1 ( ) 2 ( ) 3

In the following type of item, the respondent answers in accordance with the following instructions:

Below is a list of paired characteristics. You are to place an "X" before the item of a pair which comes closest to fitting the behavior of the average person. Be sure that you mark one of each set of statements.

( ) fears another war taking place
( ) fears doing the wrong thing

Examination of the items above shows that the subject, in giving answers, would presumably need to respond through a projective mechanism of one sort or another.

D. Title and General Description

The title of the complete measuring instrument is the *Freeman* "MA" *and* "PS" *Test*. "MA," of course, stands for Manifest Anxiety; "PS" for Psychosomatic. The "MA" test will be found in the appendix.

It will be observed that the instructions on the face sheet give the subject the impression that the purpose of the "MA" test is to determine his ability to judge the behavior of other people. The purpose of this disguise is to remove all possible tension and defensiveness in the subject in taking the test.

The "MA" part of the test contains five subsections listed as A, B, C, D, and E. There is a total of 141 items in these combined sections.

E. Technique of Administration

In an effort to standardize the "MA" test in realistic situations, the test was administered only under conditions which led the subject to regard the taking of the test as a serious performance.

Hospital and nonhospital subjects were given the test singly or in small groups. In some cases, the tests were given by the head psychiatrist or his assistant, while in other cases they were administered by the investigator, professor, or by the personnel manager. However, great care was taken in each instance to provide uniform instructions to the subjects.

In the hospital situations, the patients were told that the test was being given because the doctors needed the information to assist them in making a further diagnosis. In industrial situations, the test was administered to actual job applicants. In the college classrooms the professors of the respective classes gave the tests to their students with the instructions that the test results would be used by the professor for his own evaluation of the student.

Quite often after starting the test, the subjects would approach the examiner and ask the question: "How should I take the test, the way I think others feel or the way I feel?" In each instance, the answer was given: "It does not matter, either way you please." As indicated previously, it is assumed that the subject has no basis for response other than his own experiences.

Where subjects sought to obtain clarification of a particular question in order to be assisted in giving their best answer, they were simply again referred to the preceding instructions. No other information was provided.

It was required that the signature of each subject appear on the test.

F. Subjects

In the development and validation of the test, 921 subjects were tested. It has taken a period of three years to develop the final measuring instrument.

The investigation was concerned with two principal groups: (*a*) Control ("Normals"), and (*b*) Criterion ("Manifest Anxiety Cases").[4] The two groups will hereafter be designated as the "N" and "MA" groups.

1. *The Criterion ("MA") Group*

Considerable attention has been devoted to the matter of obtaining a valid criterion. In keeping with the objective of avoiding any possible misconception regarding the character of manifest anxiety itself, every psychiatrist and resident physician was provided with the definition of manifest anxiety. In no instance did this investigator encounter any dispute regarding the existence and nature of this psychopathological trait.

Furthermore, the physicians were advised that only those cases in which manifest anxiety was severe were to be selected, i.e., termed as "Manifest Anxiety" cases. The classification of personality disorder was not of any importance for this study. Consideration was given to one factor only: the patients selected were to be those who, on the basis of medical observation, displayed signs of manifest anxiety in accordance with the definition of that term.

To further insure the validity of the criterion, sampling was performed not in one particular hospital as is so often the case, but in several different institutions. It is believed that this procedure eliminates the weakness of one-sided diagnosis, and furthermore it would seem to effect a more representative sampling of the criterion cases.

The sources of cases in the hospitals were not limited to the neuropsychiatric wards. A large percentage of the "N" cases was selected from gastrointestinal wards. Here the procedure for selection was as follows: the chief psychiatrist and the investigator visited the general wards of the hospital and consulted with the resident physicians regarding the availability of functional cases. In each instance, the resident physicians were told that the cases desired were those patients who fit the picture of the manifest-anxiety sufferer (on the basis of the definition) and whose complaints were discovered to have no organic basis. Each case submitted by the resident physician was fully discussed and had to meet the stated requirements before being accepted for the study.

---

[4] The subgroups will be discussed in a later section.

2. *The Normal ("N") Group*

Owing to the assorted composition of this group, it is apparent that the group cannot be considered as a control, in the true sense of the term. It is conceivable that among the "normal" population there are some psychotics, prepsychotics, and severe cases of anxiety neuroses. In addition, there is question whether any one is completely free of the trait of anxiety. The control group should therefore be regarded as a group of "normals" which is to be differentiated from the "criterion" group on the grounds that the latter group displayed extreme manifest anxiety.

Particular attention was paid to the matter of representative sampling. As in the case of the criterion sampling, a variety of sources was tapped. For example, the outside or nonhospital "N" group (other groups will be discussed in the following pages) was composed of (*a*) students from several different college classes, (*b*) salesmen, and (*c*) clerical workers.

G. DESIGN OF STUDY

The design of the study embodied the following considerations and procedures:

1. All groups studied for validation of the test were matched, as far as possible, on the variables of age and education so that these two factors would not materially influence the results of the investigation.

2. The problem of measuring manifest anxiety and not other variables inherent in the situation constituted a major consideration. To compare hospital "MA" cases with the nonhospital (outside) "N" cases alone would not suffice, since the factors of hospitalization could readily affect the measurement. Hence, it became necessary to attempt control of this variable by application of the measuring instrument to a similar group of hospital cases characterized by equal length of stay in the hospital, and by organic illness without "MA" involvement. With such considerations in mind, the following subgroups were composed in order to obtain items which would effect valid differentiation:

a. Hospital "MA" cases were compared with hospital "N" cases. The first step in the standardization of the test was to develop an adequate number of item-differentiators for these two groups. However, in order to make sure that hospitalization was not exerting undue influence on the results from the "N" cases, the group comparisons described in *b* and *c* below were also made.

b. Nonhospital "N" cases were compared with the hospital "MA" cases. The purpose here was to eliminate the factor of hospitalization as an influence on the development of anxiety in the "N" group. The factor of item-identity also required consideration. Presuming that the item-differentiators secured in the first investigation were for the most part measuring "MA" and not hospitalization, it would be necessary for the same item-discriminators to effect discrimination of the groups in this second procedure.

c. Nonhospital "N" cases were compared with hospital "N" cases. The purpose here was to attempt a further control of the factor of hospitalization. If the measuring instrument is a valid one, relatively few items (ideally, none) should discriminate between these two groups.

3. Since certain groups would necessarily contain both sexes, the problem of possibly measuring sex differences rather than "MA" differences also had to be taken into account. A further step in the validational procedures, therefore, was the comparison of a new group of 100 hospital "MA" male patients with a new group of 50 nonhospital female subjects, to determine if sex difference influenced the results of the investigation.

4. A reliability study was performed upon a new group of 100 nonhospital "N" subjects.

5. The reliability study was extended to include those items which were "nearly significant," to determine if the inclusion of these items would increase the reliability of the test.

6. The design further examined the significant items to determine (*a*) which type of item was most successful in differentiating the "MA" from the non-"MA" cases; and (*b*) what implications could be derived from the content of the valid items.

7. Finally, the validity of total scores on the test was determined, and norms were established.

H. VALIDATION AND STANDARDIZATION OF THE TEST

The test was validated on a total of 921 subjects.[5] Of this number, 405 were used in preliminary investigations (160 criterion cases and 245 control subjects) of both sexes. Since a detailed report of the results of these preliminary studies would heavily overload this paper with statistical data, only occasional reference will be made to these earlier investigations in the course of the present study.

The 616 different subjects comprising the group used in the *final* procedures in the validation of the test are classifiable as follows:

Item-validation studies: Total 366
  150 criterion cases
  216 control subjects

Reliability studies: Total 100
  100 outside (nonhospital) "N" subjects

Sex-difference study: Total 150
  (Includes the 100 "N" of the reliability study)

Cross-validation of total test: Total 200
  (Includes the 100 "N" of the reliability study)

The procedure in this investigation for the validation of the test differs from the usual method of test validation. In the latter instance, standardization of a test is accomplished by accumulating a number of significant indicators and then applying the completed instrument to a new group of criterion cases for validation of total scores obtained by the instrument. In this investigation, significant items were developed in preliminary item-analysis studies, and *each item* of the completed instrument was validated by comparison of responses in various "MA" vs. "N" groups. This procedure has the advantage of validating every single item of the test, through the establishment of item-identity in all of the validity comparisons made. The fact that a total of 771 subjects was used in the preliminary and final applications of this validation procedure, of which number 310 were "MA" cases, should adequately establish the validity of the instrument *on the basis of constancy of item-differentiation.*

---

[5] The hospital "MA" and "N" cases were obtained from the White Memorial Hospital and the Cedars of Lebanon Hospital in Los Angeles; from the Los Angeles County General Hospital; and from the Veterans Administration Hospital, Van Nuys, California.

I. FINAL ITEM-VALIDATION STUDIES 1 AND 2

The purpose in Final Study No. 1 was to sample 100 "N" hospital cases and 150 "MA" hospital cases for further validation of the significant items obtained with similar groups and with similar numbers in the preliminary investigations. This validation study was confined to hospital cases only, so that the effect of hospitalization might "cancel out" in the two groups.

The groups were successfully matched for education and age distributions, there being no significant differences between the standard deviations for age and education in the two groups (*t* of .03 and *t* of 1.00).

Of 141 items, 18 showed significant phi values (10 significant at .05 level and 8 significant at .01 level). In all the investigations, both preliminary and final, Guilford's phi-coefficient was employed as a measure of the power of each item to distinguish between the groups compared.

Final Study No. 2 deals with two groups of "N" cases. One of the groups is institutional, consisting of 100 hospital "N" cases, and the other is a group of 116 outside normals.[6] The purpose here was to determine if the normals in the hospitals responded to the test items similarly to the nonhospital normals.

An attempt was made to match the two groups for age and education. The results indicate that while the groups were matched on the variable of education, a statistically significant difference exists in the variability of age (*t*'s = .03 and 2.77).

The statistical findings obtained from the comparisons of the two groups show eleven significant items.

The tentative observation which may be formed on the basis of these statistical comparisons is that the comparison of the normal groups produced fewer discriminating items than the comparisons of the hospital normals with the "MA" hospital cases. This finding is in line with the data of the preliminary studies, indicating that the items tend specially to discriminate the "MA" cases.

J. FINAL ITEM-VALIDATION STUDY 3

In this investigation, the plan was to compare a group of 116 nonhospital "N" subjects with a group of 150 "MA" hospital patients in order to obtain valid item-differentiators. The first step was to attempt to match the groups on the variables of age and education. The two groups were successfully matched in education (no significant difference between means or standard deviations, respectively). Although the groups were not successfully matched in age, the difference does not appear to hold practical significance. Item analysis of the results of the comparisons of the two groups resulted in securing 40 significant phi values.

Table 1 presents comparisons of the results obtained in the final item-analysis studies, as well as related data.

K. EVALUATION: CORRESPONDENCE OF SIGNIFICANT PHI VALUES IN FINAL STUDY 1 AND FINAL STUDY 3

In comparing the significant differentiators of Study 1 with those of Study 3, the following findings were secured:

1. Fourteen of the 18 significant values in Study 1 were also significant in Study 3, indicating that the test was differentiating "MA" cases from cases in the two "N" groups, and also suggesting

---

[6] This is a representative group composed of 21 school teachers, 17 students of a class in speech, 48 salesmen, and 30 industrial applicants.

TABLE 1

PHI COEFFICIENTS IN PRESENT STUDIES, FOR ITEMS FOUND SIGNIFICANTLY DISCRIMINATIVE IN PREVIOUS STUDIES

Final Study 1 (Males)	Final Study 2 (Males)	Final Study 3 (Males)
100 hospital "N" cases	100 hospital "N" cases	116 nonhospital "N" cases
150 "MA" hospital cases	116 nonhospital "N" cases	150 hospital "MA" cases
Significant $\phi = .121$ Very signif. $\phi = .162$	.135 .177	.120 .160
	Number of Significant Phi's	
18 (10 sig.) (8 very sig.)	11 (3 sig.) (8 very sig.)	40 (12 sig.) (28 very sig.)

that the significant items of the test are relatively immune to hospitalization influences.

2. Since 4 of the 14 items in Study 1 increased their significance from the 5 to the 1 per cent level in Study 3, this may be taken to mean that either the confinement of hospitalization itself or the presence of an organic illness develops some measure of anxiety in the organically ill, but not enough to identify them with the "MA" syndrome.

3. All of the values which were significant at the 1 per cent level in Study 1 were equally significant in Study 3. This evidence of item-identity further establishes the validity of the indicators.

4. Since in Study 3, 28 of the 40 differentiators were at the 1 per cent level of significance as against 8 of 10 in Study 1, the evidence strongly suggests the observation that the test is measuring successive degrees of anxiety, from the nonhospital normals, through the organically ill (hospitalized) normals, to the hospitalized "MA" cases.

L. FINAL STUDY 1 COMPARED WITH FINAL STUDY 2

In comparing Study 1 (hospital "N" and hospital "MA" cases) with Study 2 (hospital "N" and nonhospital "N"), the following results were secured:

1. Only 4 of the 18 significant items in Study 1 showed correspondence of significance in Study 2, thereby again indicating that only a slight increase in anxiety may be considered to exist in the hospital "N" cases over the nonhospital "N" subjects.

2. Since only 11 indicators were established in Study 2 between the two groups of "N" cases, 18 between the hospital "N" and the hospital "MA," and 40 between the nonhospital "N" and hospital "MA," the results may be regarded as showing the instrument to be a valid measure of "MA," both as to extreme (hospitalized "MA") cases and the middle of the continuum (hospital "N" cases).

M. FINAL STUDIES 2 AND 3

Comparison of results of these two studies (hospital "N" and nonhospital "N" cases, vs. "MA" hospital and nonhospital "N") gave the following results:

1. Nine of the 11 significant indicators in Study 2 are the same items as in Study 3, thereby establishing item-identity.

2. Eight of the 9 items are at the 1 per cent level of significance in both studies.

3. Study 3 gained 31 new significant

indicators when the nonhospital "N" cases were compared with the criterion cases.

N. Discriminating Items

In all of the item-analysis studies, the type of item in subsection "A" of the test typically maintained its consistency in being the most successful differentiator between the "MA" and the non-"MA" cases. Of the total number of valid items, 32 were of the "A" classification, only one significant indicator was secured from the "B" category, none from sections "C" and "D," while 7 emerged from the "E" classification.

O. Reliability Study I

This investigation was based upon a new group of 100 nonhospital "N" subjects that included: 17 salesmen in industry (who were informed that their employment would depend upon their obtaining a satisfactory score on the test), 20 fraternity applicants (who were given the inventory with the statement that their test scores would serve as a basis for evaluating their qualifications for admission into the fraternity), and 63 students in speech, drama, and commerce. The professors of the respective classes gave the tests to their male students with the instructions that the test results would be used by the professor for his own evaluation of the student.

The Kuder-Richardson formula #21 was employed for a determination of the reliability coefficient. The results of this preliminary investigation are as follows: Range 20, Mean 17.04, *SD* 3.94, and Reliability Coefficient, .49.

Since the reliability coefficient was not high,[7] the K-R formula was applied in another preliminary study with weighted scores.[8] The results showed that the coefficient based on weights turned out to be essentially the same ($r_{11} = .48$).

P. Reliability Study II

Procedure

The design in this reliability study was to determine whether increasing the number of item-differentiators would raise the reliability of the test. Accordingly, a procedure was followed in which all of the phi coefficients were placed in rank order, and the 56 items which had phi values of .09 or over were selected.[9]

The results of this study indicate that including the "nearly significant" phi's raises the reliability coefficient from .49 in the first reliability study to the relatively high level of .73 in this present

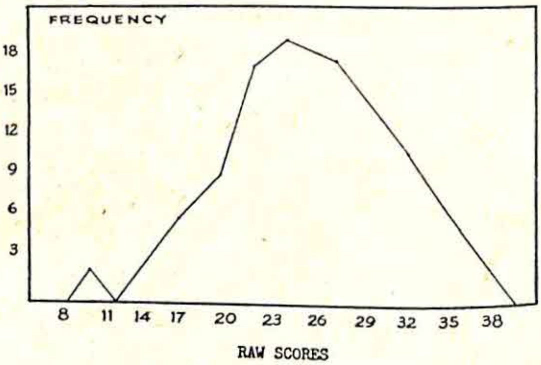


FIG. 1. Distribution of Raw Scores Based on 56 Items: Final Reliability Study, 100 Outside "N" Cases.

---

[7] Guilford (8) in discussing the question of reliability has stated that the old standards for reliability and validity need serious re-evaluation. He cites three dramatic instances where three useful tests whose statistics were based on large samples had reliabilities below .40.

[8] Guilford (6) recommends a simple approximation weight for general use and provides an *abac* for its estimation when the correlation between item and the criterion is the *phi* coefficient. He derived a formula for the standard error of this weight and presents tables of significant and very significant weights in terms of deviations from the median weight.

[9] The minimal value of .12 was used in the preceding studies. This "N" group represents the same group of 100 "N" cases used in the preliminary reliability study.

TABLE 2

COMPARISON OF TOTAL SCORES OF 100 HOSPITAL "MA" AND 100 NONHOSPITAL NORMAL CASES

(Total Scores Based on 56 Items)

Group	Range of Scores	Mean	*SD*
100 "MA" hospital cases	30	29.30	7.29
100 "N" nonhospital cases	20	24.81	6.18
Critical score: 28			

Tests of Significance

t of Means	3.90
Validity Coefficient ϕ	32
Equivalent of Pearson r (r_ϕ)	.40*
Chi Square	13.31**

* Since the scores on the tests represent a continuous distribution and the two groups of "N" and "MA" act as a dichotomy and may be considered a point distribution, phi becomes equivalent to a Pearson r by the formula $r_\phi = \frac{\phi}{.798}$.

** With one degree of freedom, a chi square of 6.635 indicates significance at the 1 per cent level.

study. Range of scores was 29; the mean score became 24.81; the *SD,* 6.18.

Figure 1 provides a graphic distribution of the scores. It is to be observed that the distribution approximates a normal curve.[10]

Q. STUDY OF SEX DIFFERENCE AS A FACTOR IN ANXIETY

The investigation here was concerned with two new groups of nonhospital "N" subjects (100 males and 50 females) selected at random from college students, to determine if the variable of sex difference is a contributing factor to anxiety.

On the basis of the results obtained in this study, it may be stated that no significant difference exists between the scores of the two groups. Sex does not seem to be a determining factor in the formation of anxiety (t of means = .89).

R. VALIDATION OF TEST: TOTAL SCORES

The purpose of this study was to validate the total scores of the test, using new groups of outside "N" and "MA" cases for the development of a scoring key. The procedures employed were as follows:

1. A new group of 100 "MA" hospital cases was given the test.[11]
2. The scores were compared with the scores of the 100 nonhospital "N" cases which had been used in the previous reliability study.
3. Fisher's t test of significance was applied to the means of the test scores.
4. A cut-off score was established by the graphic method.
5. The phi coefficient and chi square were employed for validity studies.

Table 2 presents relevant data.

1. *t of Means*

The data in Table 2 show a very sig-

---

[10] The 56 items with significant or "nearly significant" phi values are identified in the "MA" test in the appendix.

[11] The same procedures were followed by the psychiatrists in the selection of this new group of 100 "MA" cases as characterized the previous psychiatric samplings of the hospital "MA" cases.

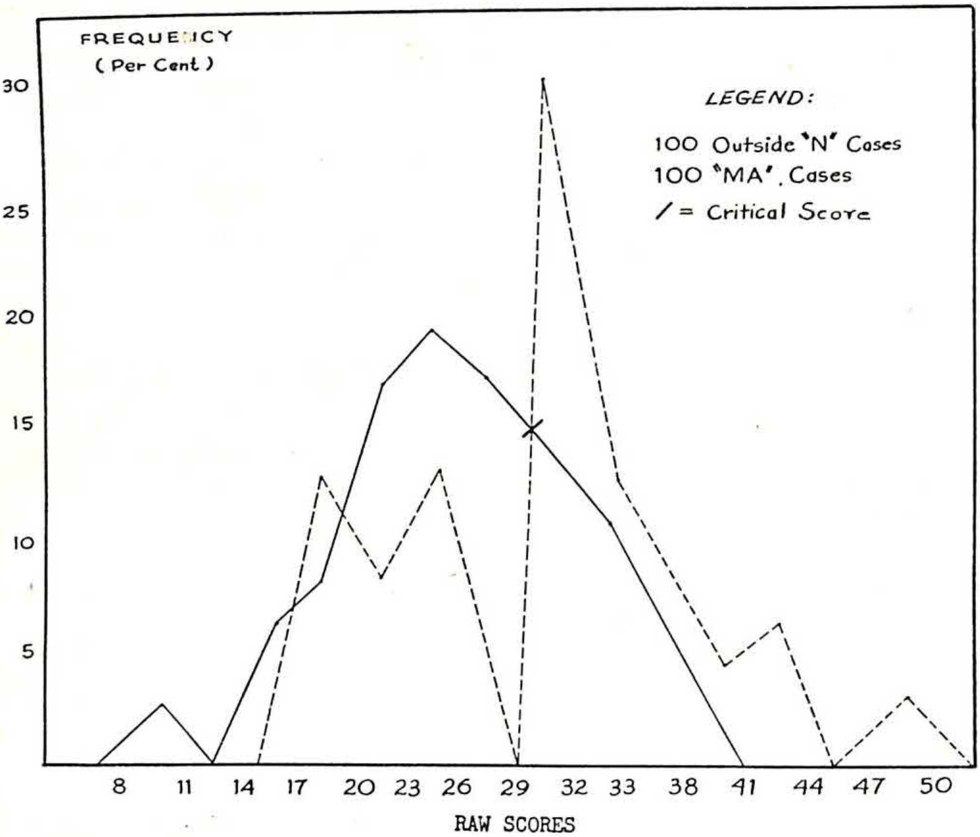


FIG. 2. Distribution of Raw Scores Based on 56 Items: Validation Study, 100 Outside "N" Cases and 100 "MA" Cases.

nificant difference in the mean scores when the "N" group is compared with the "MA" group. This would indicate that there is less than one chance in a hundred that the observed mean difference between "N" and "MA" scores could be produced by chance. The null hypothesis can therefore be rejected with confidence.

Figure 2 shows the distributions of the scores.

2. *Critical score*

The cut-off score was established by the graphic method of locating the point of intersection of the curves of the distributions of the two groups. The critical score is 28.

3. *Validity coefficient "phi" and value of "chi square"*

The validity coefficient, Φ, equals .32; in terms of equivalent Pearson *r*, this equals .40. For a significance-test, phi was converted into chi square (7, p. 246).[12] As indicated in Table 2, chi square equals 13.31. Since the required figure for significance at the 1 per cent level is 6.635, there is less than one chance in a hundred that the obtained chi square could occur by chance. The result is therefore very significant and the null hypothesis can be rejected with confidence.

S. THE CRITICAL SCORE

The critical score used in separating normals from abnormals cannot be considered as a clear-cut dividing line. The problem of such separation has always been a difficult matter which has elicited much discussion (2). The term "MA" has been defined in this study as amounts of anxiety observed to be present in a very severe degree. Yet "normals" who possess anxiety in less amounts (below the critical score) may also show disturbed behavior.

---

[12] The formula reads $\chi^2 = N\Phi^2$. Guilford states that the standard error for phi is too laborious to be of practical value and he recommends that, for tests of significance, chi square, derived from phi, be employed.

The clinical setting, however, demands some dividing line in order to assist in the identification of the "MA" case. To this extent, a critical score becomes valuable.

TABLE 3

"MA" Centile Norms Based on 100 Nonhospital "N" Cases

Centile	Score Point	Integral Score
99	38.5	39
95	34.5	35
90	33.0	33
85	32.0	32
80	30.0	30
75	29.0	29
70 (c.s.)*	28.2	28
60	26.7	27
50	25.0	25
40	23.3	23
30	21.5	22
20	19.8	20
10	16.6	17

* c.s. = cut-off score.

In order to complete the development of the scoring key, a percentile table has been constructed, as shown in Table 3. The cut-off score falls at the 70th percentile of nonhospital "N" cases, which means that 30 per cent of the nonhospital "N" population have scores higher than the cut-off scores presented by the test.

T. Conclusion and Implications

The preliminary item-analysis studies, the three final item-validation studies, the two reliability studies, the further study of sex difference, and the final validation study for the establishment of a cutting score and norms seem to justify the following conclusions:

1. A pencil-and-paper personality test can become a reasonably valid measuring instrument if structured on a projective-identification basis such as the "MA" Test developed in this investigation. The "MA" Test maintains the advantage over other objective scoring instruments through (*a*) its feature of disguise, the test being ostensibly designed to evaluate the subject's ability to judge (socially) the behavior of other people; (*b*) elimination of conscious insight into the character and intent of the items, thereby reducing the factor of "faking" and manipulation of the test items; and (*c*) causing the respondent to project his unconscious identifications for revelation of his personal anxiety characteristics. Inspection of responses to the individual items of the "MA" Test also allows the examiner to locate the significant indicators of manifest anxiety and to estimate the nature and extent of the subject's manifestations of anxiety.

2. The implications resulting from the above findings should prove to be highly significant for clinical psychology. If, as seems likely, anxiety reflects a fundamental, nuclear manifestation in personality disorders, it would seem that future clinical research should be directed towards a more adequate understanding of the effects of fear upon the development of emotional pathology and biofunctional disturbances. Similarly, from the standpoint of psychotherapy, effective reconstruction of the personality would need to take into account adequate modification or eradication of the nuclear pathological elements of fear and anxiety.

3. The securing of a low score on the

"MA" Test should not be regarded as an indicator of absence of anxiety or presence of good emotional adjustment. Empirically, the results suggest the conclusion that scores below 1 *SD* from the mean may be found to characterize either the psychopaths or the anxiety cases who have developed a "comfort trend" in their behavior. In both these instances, the neurosis manifestations would involve a pathological deficit of normal apprehension and conflict through "blotting out" reality, or through relatively uninhibited libidinal expression.

4. Of particular significance is the relationship of the "MA" (manifest anxiety) nucleus to the various pathological directions of the neurosis disorder, including the intellectual functioning. It is suggested that future research could take into consideration possible relationships between high "MA" scores and clinically suggestive signs of response patterning on the Wechsler-Bellevue test.

The results of this investigation point in the direction of anxiety being nuclear to various forms of neurosis. The basis for this statement is the fact that all forms of neurotic disorder are represented in the group of criterion cases, who were selected only for "manifest anxiety" (see pages 3-4).[13]

---

[13] A subsequent paper based on further research suggests the hypothesis that manifest anxiety is nuclear to the psychoneuroses in general. If still further research bears this out, the significance of the "MA" test will obviously be greatly enhanced.

APPENDIX

THE FREEMAN MA TEST[1]

Name .... Age .... Sex ....

Address .... Occupation ....

Date of Test .... Status ....

INSTRUCTIONS

The purpose of this test is to tell how well you are able to judge the behavior of other people. In some questions you are asked merely to check "yes" or "no." In others you are to choose the right answer. Read each question carefully, but do not spend too much time on any one question.

EXAMPLE

Questions	*Answers*
1. A boastful person has a weak character	Yes ( ) No (×)
2. A very fussy person is (a) very hard to get along with (b) no trouble at all	a (×) b ( )

"A"

QUESTIONS	ANSWERS[2]	Phi[3]
1. When someone says the wrong thing in conversation, he is bothered by it afterwards.	Yes (×) No ( )	.17
2. One who often loses his temper feels sorry afterwards.	Yes (×) No ( )	.35
3. If someone is easily irritated, the reason is he cannot help himself.	Yes (×) No ( )	.18

---

[1] Copyright, 1952.

[2] "×" = response scored positively as an indicator of manifest anxiety ("MA").

[3] Only values of phi equal to .09 or greater are listed. Most of these items are statistically significant at the .05 level or better. (Cf. Table 1.)

QUESTIONS	ANSWERS	Phi
4. If a person worries a great deal about his job, the reason is that he is afraid of losing his job.	Yes (×) No ( )	.13
5. When a person realizes that he has made a bad mistake, he usually condemns himself for having made that mistake.	Yes ( ) No ( )	...
6. Most people believe that a murderer should get the electric chair.	Yes ( ) No ( )	...
7. The average person becomes "weak" at the sight of blood.	Yes (×) No ( )	.16
8. Looking down from a high building is usually frightening.	Yes (×) No ( )	.26
9. It is a very good thing that most people go to the doctor at the first sign of illness or pain.	Yes (×) No ( )	.14
10. Most people are unable to appreciate how miserable it is to have trouble sleeping at night.	Yes (×) No ( )	.19
11. Going to the blood bank for the first time usually causes a person to be fearful.	Yes (×) No ( )	.19
12. A good driver is one who always has on his mind the many possible accidents that can happen to him.	Yes (×) No ( )	.17
13. Usually when there is pain in one part of the body, other parts of the body also become affected with pain.	Yes (×) No ( )	.10
14. The conscientious person continually keeps thinking of all the other things that have to be done before the day's work is finished.	Yes (×) No ( )	.25
15. In order to be successful, a person should always be worried about doing the wrong thing.	Yes (×) No ( )	.21
16. If a person is very sensitive, the reason is that he has a fine character.	Yes (×) No ( )	.20
17. From the standpoint of desirable health standards, most people are lacking in sufficient cleanliness.	Yes (×) No ( )	.34
18. The person who is exceptionally concerned about his future security is a practical person.	Yes (×) No ( )	.18
19. The person who always has trouble making a decision wants to be sure he is not going to make a mistake.	Yes (×) No ( )	.28
20. Being in big crowds of people will frequently cause a person to experience a feeling of suffocation.	Yes (×) No ( )	.47
21. A sensible person always keeps the window blinds down at night to avoid being seen.	Yes (×) No ( )	.33
22. People who have very few friends are likely to be very selfish people.	Yes (×) No ( )	.21
23. A person who has a great deal of energy is usually a very nervous person.	Yes (×) No ( )	.15
24. An honest person will very seldom become a politician.	Yes ( ) No ( )	...
25. The person who has a great many fears is usually a coward.	Yes (×) No ( )	.14
26. People who reach the top of the ladder in industry usually get there because of some pull.	Yes (×) No ( )	.26
27. Many people suffer from extreme fatigue.	Yes ( ) No ( )	...
28. Going to the movies is one of the best ways to overcome discouragement.	Yes (×) No ( )	.19
29. The difficulty with most relatives is that they are usually very bothersome and jealous.	Yes (×) No ( )	.27
30. There are very few people who really understand us well.	Yes ( ) No ( )	...
31. It is correct to say that the average person wastes much of his time.	Yes ( ) No ( )	...
32. Most people find it particularly annoying to have to wait for a friend who is late for his appointment.	Yes (×) No ( )	.11
33. Most of us live under a "nervous tension" because of our "city way" of living.	Yes (×) No ( )	.26
34. Usually it is best not to trust anything a salesman says of his product.	Yes (×) No ( )	.29
35. The average person finds himself putting heart and soul into his work.	Yes ( ) No ( )	...
36. The average person very often checks his work to make sure that he will not make a mistake.	Yes ( ) No ( )	...
37. In handling daily duties, one cannot be at ease until the very last thing is out of the way.	Yes (×) No ( )	.16
38. As a rule, one does not tell a person about his faults because one does not wish to hurt his feelings.	Yes ( ) No ( )	...
39. When something of importance is about to take place, a person usually thinks of the many things that may go wrong.	Yes ( ) No ( )	...

QUESTIONS	ANSWERS	Phi
40. The trouble with most people is that they do not pay enough attention to turning out perfect work.	Yes ( ) No ( )	...
41. Very few people are sufficiently tidy about their work.	Yes (×) No ( )	.11
42. Only some people have the exceptional ability to judge whether a person is telling the truth.	Yes (×) No ( )	.14
43. It is usually very upsetting to have to read in the newspapers about a terrible tragedy that has taken place.	Yes ( ) No ( )	...
44. It is best to have only a few friends, but good ones.	Yes (×) No ( )	.22
45. One typically finds that the experiences which were at first feared turned out to have no danger to them at all.	Yes ( ) No ( )	...
46. Most people find it disturbing to read of the many murders and accidents that are reported in the newspapers.	Yes (×) No ( )	.23
47. As a rule people's fears concern small silly things rather than large events.	Yes ( ) No ( )	...
48. In this day and age it is very difficult to find enough time in which to relax.	Yes (×) No ( )	.20
49. One has always so many things to do that it becomes difficult to avoid being rushed.	Yes (×) No ( )	.19
50. There are more dishonest than honest people in the world.	Yes (×) No ( )	.30
51. A conscientious person is one who spends a lot of time thinking about his problems.	Yes (×) No ( )	.23

"B" (CHOOSE 1 or 2)

52. The person who presents the more undesirable behavior is (1) the one who criticizes people (2) the person who lies occasionally	1 (×) 2 ( )	.09
53. The less bothersome person would be (1) the one who argues a lot (2) the one who has a weak character	1 ( ) 2 ( )	...
54. Someone has trouble making a decision. The reason is that (1) he is giving careful thought to both sides of the question (2) he doesn't trust his judgment	1 (×) 2 ( )	.13
55. The one who would prove to be most undesirable as a friend would be (1) the person who is very stubborn (2) the person who lacks education	1 (×) 2 ( )	.10
56. A person would suffer more from (1) nervousness (2) rheumatism	1 ( ) 2 ( )	...
57. The person who works rapidly at his job is (1) a nervous worker (2) an efficient person	1 ( ) 2 ( )	...
58. A person is the life of the party because (1) he is always happy (2) he is usually unhappy	1 (×) 2 ( )	.12
59. The successful person (1) takes things as they come (2) worries before each task	1 (×) 2 ( )	.11
60. A person with good character feels sorry for people in trouble (1) once in a while (2) all the time	1 ( ) 2 ( )	...
61. People are usually more bothered by (1) loud noises (2) headaches	1 ( ) 2 ( )	...
62. People are usually more aggravated by (1) being disappointed about something (2) not being able to solve problems	1 ( ) 2 ( )	...
63. The average person attempts to enjoy himself by (1) usually doing something (2) usually relaxing	1 ( ) 2 ( )	...

QUESTIONS	ANSWERS	Phi
64. As a rule, when something disturbing takes place (1) it does not last long (2) usually causes a person to take it to heart	1 ( ) 2 ( )	...
65. Of the two, the more objectionable person is the one (1) who is dishonest at times (2) who talks too much	1 ( ) 2 ( )	...
66. A person is constantly active because (1) he does not like to waste his time (2) he finds it hard to do nothing	1 (×) 2 ( )	.27
67. A person who wants to satisfy his wishes without much delay is (1) an impulsive person (2) a very successful person	1 ( ) 2 ( )	...
68. John Doe has these two fears. He is more disturbed by (1) fear of death through an automobile accident (2) fear of dying during the night	1 (×) 2 ( )	.09
69. Of the two, a person is more embarrassed when (1) realizing he was lost in thought while someone was talking to him (2) dropping his fork at the dinner table	1 ( ) 2 ( )	...
70. People as a rule suffer mostly from (1) general aches and pains (2) one specific illness	1 ( ) 2 ( )	...
71. Forgetfulness usually takes place (1) at work (2) at occasions when someone's name has to be recalled	1 ( ) 2 ( )	...
72. (1) most people are very conscientious (2) few people are very conscientious	1 ( ) 2 ( )	...
73. People are more frequently disturbed by (1) looking down from a tall building (2) not having sufficient air in the room	1 ( ) 2 ( )	...
"C" (CHOOSE 1, 2, or 3)		
74. When a supervisor is too "bossy" the employee feels that (1) the supervisor should be "fired" (2) should be lowered to the employee's rank (3) should not pay attention to the supervisor	1 ( ) 2 ( ) 3 ( )	...
75. A person who prefers to be alone is usually (1) an unfriendly person (2) likes to mind his own business (3) a home lover	1 ( ) 2 ( ) 3 ( )	...
76. As far as imagination is concerned, most people have (1) just an ordinary amount (2) a small amount (3) a great amount	1 ( ) 2 ( ) 3 ( )	...
77. One usually finds that the average person regards his personal belongings with (1) ordinary care (2) very careful attention (3) not enough care	1 ( ) 2 ( ) 3 ( )	...
78. A person usually worries about things because (1) he can't help himself (2) he gives serious thought to his problems (3) tries to avoid doing the wrong thing	1 ( ) 2 ( ) 3 ( )	...
79. One of your friends is bothered by these three things. The one that aggravates him the most is (1) shyness (2) poverty (3) irritability	1 ( ) 2 ( ) 3 ( )	...
80. Another one of your friends suffers from the following. The most bothersome one is		

QUESTIONS ANSWERS Phi

(1) sensitiveness
(2) rheumatism
(3) nervousness 1 ( ) 2 ( ) 3 ( ) ...

81. The ideal employee would be one who is
(1) exceptionally careful about details
(2) ordinarily careful
(3) doesn't waste too much time about details 1 ( ) 2 ( ) 3 ( ) ...

82. Judging our friends, we should say that most of them are
(1) exceptionally happy
(2) occasionally happy
(3) never really happy 1 ( ) 2 ( ) 3 ( ) ...

83. In trying to be successful at work, one should
(1) stay overtime and try to get everything done
(2) try not to become discouraged
(3) work conscientiously 1 ( ) 2 ( ) 3 ( ) ...

84. John Doe has three problems. He will first try to overcome
(1) being fearful
(2) being in debt
(3) lack of education 1 ( ) 2 ( ) 3 ( ) ...

"D" (CHOOSE ONE)

85. Someone you know shows these undesirable traits. The most aggravating trait that he should first start to overcome is
(1) telling lies
(2) not being on time
(3) laziness 1 ( ) 2 ( ) ...
(4) nervousness 3 ( ) 4 ( ) ...

86. A person who has all of the following ailments is probably mostly upset by
(1) inability to sleep well
(2) continual head colds
(3) upset stomach 1 ( ) 2 ( ) ...
(4) earache 3 ( ) 4 ( ) ...

87. Of the following four traits, the most undesirable one is
(1) a person who continually criticizes
(2) who is domineering
(3) who shows dishonesty 1 ( ) 2 ( ) ...
(4) who does not save any money 3 ( ) 4 ( ) ...

88. If one of your friends showed the following traits, the most disturbing one would be
(1) his being superstitious
(2) his becoming easily irritable
(3) having feelings of insecurity 1 ( ) 2 ( ) ...
(4) having many prejudices 3 ( ) 4 ( ) ...

89. A person hurts himself mostly from
(1) working too much
(2) frequently losing his temper
(3) too much anxiety 1 ( ) 2 ( ) ...
(4) not saving for a "rainy day" 3 ( ) 4 ( ) ...

"E"

Below is a list of paired characteristics. You are to place an "X" before each one of the pairs which comes the closest to fitting the behavior of the average person.

Here is an example
( ) Forgetfulness
(X) Does not read enough

Be sure that you mark one of each set of statements.

Phi

... 90. ( ) Tendency to be lost in thought
( ) Does not read enough

Phi		
...	91.	( ) Becomes discouraged at times
		( ) Irritable at times
...	92.	( ) Inclined towards worry
		( ) Inclined towards fast driving
...	93.	( ) Worries about work
		( ) Does not belong to enough organizations
...	94.	( ) Has few friends
		( ) Makes mistakes occasionally
.15	95.	(×) Fears looking down from a tall building
		( ) Fears riding in an elevator
.11	96.	(×) Fears losing his job
		( ) Fears making a mistake
...	97.	( ) Feels miserable at times
		( ) Feels lazy at times
...	98.	( ) Often severely criticizes himself for doing the wrong thing
		( ) Often forgets to remember peoples' names after first meeting them
.09	99.	(×) Not satisfied until every detail is properly handled
		( ) Not satisfied until objectives are reached
.21	100.	(×) Fears another war taking place
		( ) Fears doing the wrong thing
...	101.	( ) Tendency to postpone doing things
		( ) Does not take enough time for lunch
...	102.	( ) Tendency to overlook details
		( ) Dislikes riding street cars
...	103.	( ) Too exacting in work
		( ) Occasional carelessness
...	104.	( ) Dislikes being bossed
		( ) Dislikes heavy traffic
...	105.	( ) Keeps hurt feelings to himself
		( ) Occasional loss of temper
...	106.	( ) Tries to keep from being aggravated
		( ) Tries not to be late
...	107.	( ) Finds it difficult to make new friends
		( ) Finds it difficult to save enough money
...	108.	( ) Tries hard to be nice to people
		( ) Tries to be well informed on all subjects
...	109.	( ) Tendency to look at bad side of future happenings
		( ) A dislike for getting into an argument
.09	110.	(×) Has been called stubborn by others
		( ) Has been criticized for being late for an appointment
...	111.	( ) Has been told that he is forgetful
		( ) Has been criticized for being late for an appointment
.10	112.	(×) Often cannot overcome certain problems
		( ) Often changes ideas about things
.13	113.	(×) Wishes people were more tactful
		( ) Wishes people were more sensible
...	114.	( ) Wishes he had more time in which to do things
		( ) Wishes people would be more cooperative
...	115.	( ) Often regrets saying things
		( ) Often likes to sleep later in the morning
.13	116.	(×) Has habit of being too careful in making a decision
		( ) Has habit of not giving enough thought to old age security.
.13	117.	(×) Has tendency to feel bad when things do not work out as they should
		( ) Has tendency to forget to remember a relative's birthday.
...	118.	( ) Likes to attend movies frequently
		( ) Likes to play cards with others
...	119.	( ) Believes in leading a practical life
		( ) Believes in good neighbor policy
.18	120.	(×) Always active and on the go

Phi
( ) Always interested in enjoying life.
... 121. ( ) Dislikes being idle
( ) Dislikes Communism
... 122. ( ) Eager to know what the next day will bring
( ) Eager to know the news of the world
... 123. ( ) Wishes there were less unhappiness in this world
( ) Wishes parking facilities were better
... 124. ( ) Believes nervousness can be overcome
( ) Believes nervousness is very difficult to overcome
... 125. ( ) Wants to get rid of all personal fears
( ) Wants to get rid of all unfinished tasks
... 126. ( ) Would like to see Capital and Labor get along
( ) Would like to have people be more friendly
... 127. ( ) Would like to see more education for everyone
( ) Would like less politics in government
... 128. ( ) At times has trouble making a decision
( ) At times feels a need for rest
... 129. ( ) Would wish people would drive more carefully
( ) Would wish a work week would consist of three days
... 130. ( ) Would regard lack of patience as being a form of nervousness
( ) Would say lack of patience is due to over-anxiousness in wanting to get a thing done
... 131. ( ) Would like to enjoy a good night's sleep
( ) Would like to take a vacation
... 132. ( ) Would always like to have a healthy heart
( ) Would always like to be happy
.15 133. (×) Would like to see a better world in which to live
( ) Would like to see a reduction of taxes
... 134. ( ) Believes people need to understand him better
( ) Believes people need to develop more education
... 135. ( ) Most people are careless in doing their work
( ) Most people need more hobbies
... 136. ( ) The average person works leisurely
( ) The average person finds a day's work is usually tiring
... 137. ( ) The average employer fully fails to appreciate his employee
( ) The average employer doesn't take time to learn what his employee accomplishes
.10 138. (×) Wishes people's criticisms wouldn't bother him so much
( ) Wishes industry were capable of more production
... 139. ( ) Finds that when he hates someone it doesn't last long
( ) Many people could develop better judgment
... 140. ( ) Whenever a person is restless he seeks to overcome it
( ) Whenever someone has cheated, he is distrusted thereafter
... 141. ( ) Finds that usually when he fears something is going to happen, it does not take place
( ) Finds that he dislikes careless drivers.

REFERENCES

1. Babcock, H. A. A neurosis? Or neurotic behavior? *J. Psychol.*, 1944, **17**, 61-74.
2. Ellis, A. The validity of personality questionnaires. *Psychol. Bull.*, 1946, **43**, 385-441.
3. Freeman, M. J. The standardization of a psychosomatic test: validation of a psychosomatic inventory. *J. Pers.*, 1950, **19**, 229-244.
4. Guilford, J. P. *Psychometric methods.* New York: McGraw-Hill, 1936.
5. Guilford, J. P. The phi co-efficient and chi square as indices of item validity. *Psychometrika,* 1941, **6**, 11-19.
6. Guilford, J. P. A simple scoring weight for test items and its reliability. *Psychometrika,* 1941, **6**, 367-374.
7. Guilford, J. P. *Fundamental statistics in psychology and education.* New York: McGraw-Hill, 1942.
8. Guilford, J. P. New standards for test evaluation. *Educ. psychol. Measmt,* 1946, **4**, 427-438.
9. Guilford, J. P., & Guilford, Ruth B. An analysis of the factors in a typical test of introversion-extroversion. *J. abnorm. soc. Psychol.*, 1934, **28**, 377-399.
10. Kuder, G. F., & Richardson, M. W. The theory of the estimation of test reliability. *Psychometrika,* 1937, **2**, 151-160.
11. Kuznets, G. An analysis of Bernreuter's personality inventory. *Psychol. Bull.*, 1934, **31**, 585.
12. Landis, C. Questionnaires and the study of personality. *J. nerv. ment. Dis.*, 1936, **83**, 125-134.
13. Landis, C., & Katz, S. E. The validity of certain questions which purport to measure neurotic tendencies. *J. appl. Psychol.*, 1934, **18**, 343-356.
14. Landis, C., Zubin, J., & Katz, S. E. Empirical evaluation of three personality adjustment inventories. *J. educ. Psychol.*, 1935, **26**, 321-330.
15. Lorge, I. Personality traits by fiat, I and II. *J. educ. Psychol.*, 1935, **26**, 273-278; 427-434; 652-654.
16. Malamud, D. C. Objective measurement of clinical status in psychopathological research. *Psychol. Bull.*, 1943, **3**, 240-259.
17. May, R. *The meaning of anxiety.* New York: Ronald, 1950.
18. Meehl, P. An investigation of a general normality or control factor in personality testing. *Psychol. Monogr.*, 1945, **59**, No. 4 (Whole No. 274).
19. Rulon, P. J. A simplified procedure for determining the reliability of a test by split-half. *Harvard educ. Rev.*, 1939, **9**, 99-103.
20. Sullivan, H. S. *Conceptions of modern psychiatry.* Washington, D.C.: William Alanson White Psychiatric Foundation, 1947.
21. White, R. W. *The abnormal personality.* New York: Ronald, 1948.

(Accepted for publication September 25, 1952.)

Vol. 67, No. 4 Whole No. 354, 1953

Psychological Monographs: General and Applied

Traumatic Avoidance Learning: Acquisition in Normal Dogs[1]

Richard L. Solomon and Lyman C. Wynne
Harvard University

When a "typical," normal, mongrel dog is put into a small compartment with a steel grid floor, he will usually exhibit persistent exploratory behavior. Thus confined, he will sniff at the grid bars and the corners of the compartment, walk back and forth, intermittently stand up on his hind legs with forepaws against the walls of the compartment, occasionally look over the head-high barrier which separates him from another compartment, but will make no attempt to jump over the barrier into the other compartment. If this dog is suddenly stimulated through the steel grid floor by a high-voltage electric shock, of an intensity just below the tetanizing level, the dog will immediately exhibit an intense fear reaction. If the shock is left on during the time the dog remains in the compartment, this intense fear reaction will typically contain these components: The dog will scramble rapidly and vigorously around the compartment, slamming into walls, perhaps, or leaping up against them; he will simultaneously emit a high-pitched screech, will salivate profusely, will urinate and defecate in a manner which could be called "projectile elimination," and will roll his eyes rapidly and jerkily; in addition, his pupils will dilate, portions of his hair will stand on end, small muscle groups all over his body will tremble, and his breathing will consist of short, irregular gasps. Sooner or later the dog's vigorous scrambling movements result in his getting over the barrier into the other compartment, to safety. If the shock is repeated at a later time, the same fear responses will occur again.

When a dog exhibits the components of a massive fear reaction in the presence of noxious stimulation that is very intense, we should like to define the whole event as a "traumatic experience." Such an event will typically include an intense, unconditioned stimulus which elicits an intense, unconditioned reaction pattern containing the following components: (*a*) visceral responses of high magnitude, with their correlated feedback stimulation; (*b*) skeletal responses of vigorous intensity, with their correlated proprioceptive feedback stimulation; (*c*) hormonal responses from endocrine glands and nonendocrine glands, with their correlated effects upon the body chemistry, muscles, and neural centers; and (*d*) higher brain-center activity.

One often refers to the visceral responses as emotional reactions; they are often considered to be "expressive" rather than instrumental in nature. The skeletal responses are partly expressive (elicited directly by the unconditioned stimulus) and partly instrumental; we usually interpret the instrumental portion of the skeletal reactions as "adaptive" escape or aversive movements.[2] (There is also a possibility that certain visceral responses can become instrumental, as in Yogi control of autonomic responses.)

---

[1] During the pilot stage this research was supported by a grant from the Laboratory of Social Relations, Harvard University. Thereafter, the research was supported through the generosity of the Medical Sciences Division of the Rockefeller Foundation. The authors wish to thank the following psychologists for their assistance in carrying out the experiment: Dr. Barbara N. Cohn, Dr. James Olds, Mr. Nathaniel Kogan, Mr. Leon Kamin, Mr. Martin Coles, Miss Mary Roseborough, and Miss Elaine Smulekoff. Dr. Wynne was a USPH Postdoctoral Research Fellow in mental health when this experiment was carried out.

Within the pattern of the reactions of the typical dog, when he is stimulated with an intense electric shock, are contained the essential components of an unconditioned fear reaction pattern. Because the components have all occurred with high response amplitudes, and because we know the intensity of the unconditioned stimulus to be high, we shall define the *stimulus-response pattern* as indicative of a *traumatic* experience; and any learning which has as one of its conditions the occurrence of such a stimulus-response pattern, we shall define as *traumatic learning*. Furthermore, if the traumatic learning is of such a nature that it prevents the unconditioned stimulus from occurring, we shall define it as *traumatic avoidance learning*.

We have taken pains with this definition because there is some reason to believe that traumatic learning may possibly have certain functional properties which differ qualitatively and quantitatively from those of "ordinary motivated learning."

This paper will be restricted to the characteristics of the *acquisition* of traumatic avoidance learning in dogs. Elsewhere we describe the characteristics of extinction and the effects of special procedures (18, 19).

APPARATUS

A schematic drawing of the apparatus is presented in Figure 1. It is an adaptation of the Miller-Mowrer shuttle box (11) which has been used for avoidance learning in rats. We have modified the shuttle box extensively for use with dogs.

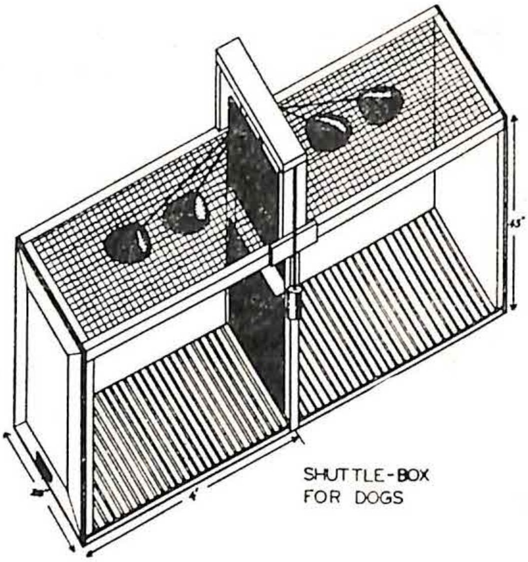


FIG. 1. Schematic drawing of the shuttle-box apparatus.

The shuttle box consisted of two compartments separated by a barrier and guillotine-type gate. The inside dimensions of each compartment were 45 in. long, 24 in. wide, and 40 in. high. The barrier was adjustable in height. The gate could be raised by means of a pulley system above the level of the top of the compartments. When the gate was closed it rested on top of the barrier, thus creating a solid wall completely separating the two compartments.

Owing to the vigorous battering the apparatus received, the construction had to be rugged. The frame was built of 2 × 4-in. pine. The walls were composed of 3/8-in. plywood, sheathed on the inner surface by 1/32-in. aluminum, half-hard plate. These aluminum inner walls were painted flat black. The doors to the two compartments are omitted in Figure 1. They were composed of 5/8-in. plywood sheathed with aluminum, painted flat black. They were hinged on the end post of each compartment and were locked in the closed position by means of slide bolts.

---

[2] There are alternative classifications in wide use today which can be *partially* substituted for the "visceral" and "skeletal" categories. For example, autonomic nervous system and central nervous system functions, involuntary and voluntary actions are often distinguished. Any of these categorizations should be used with the realization that these functions overlap considerably in the intact organism.

The ceiling of each compartment was covered with 1-in. hardware cloth (No. 8 gauge). Directly above each compartment were two 40-watt lights, shielded so that each pair cast light only into the compartment directly below. When the gate was up and only one pair of lights was on, one compartment was dark and the other illuminated. Above the gate was a wooden scaffolding on which were mounted two pulleys. A cord was strung from the middle of the top of the gate through the pulleys to a counterweight hanging at the side between the two doors. Draped over the scaffolding and the whole apparatus was a double thickness of cheesecloth serving as a crude, one-way screen.

The floor of the apparatus consisted of polished 1-in. stainless steel (channel) bars ½ in. apart, parallel to the barrier. The bars were suspended on two bars of bakelite (1 in. × ¾ in.) running the length of the apparatus outside of the compartments. The walls were supported by angle irons so that there was a ½-in. space between the bottoms of the walls and doors and the surface of the steel bars. Thus the steel bars were only in contact with bakelite outside the compartments. This type of construction insured that urine would not create an electrical contact between the bars.

The steel grid bars of each compartment were individually wired to two cannon plug receptacles, one mounted at each end of the apparatus. A multiple-strand cable connected each plug to a commutator circuit. This type of circuit for supplying fairly stable electric shocks has been described by Skinner (17). It is so designed that an alternating current is supplied to the grid bars with a rapidly shifting polarity pattern. Such a technique for distributing electric current to the grid bars effectively prevents dogs from escaping shock by shorting out adjacent grid bars with feces. The shock circuit was powered by a 500-v. A.C. transformer energized by 110-v. A.C. A fixed resistance of 3,500 ohms was placed in series with the grid bars so that the experimenters could not accidentally electrocute a dog. In addition, a 1000-ohm variable voltage divider was placed in parallel with the 500-v. power supply in such a way that the voltage at its source could be varied from 0 to 500 v. An electronic voltmeter was placed across the fixed resistance so that the voltage drop could be continuously read. In this way, the current flow in the circuit could always be known.

SUBJECTS

The subjects used in this study were 30 mongrel dogs of medium size. Their range in weight was approximately from 9 to 13 kg. All the subjects were strays and probably free from experimental sophistication. The data in this study were obtained while the dogs were apparently healthy.

PROCEDURE

The barrier was adjusted to approximately the height of the dog's back. Therefore, the dogs could look over the barrier from one compartment to the other when the gate was raised.

Pretesting. Each dog was pretested in order to reveal the presence of a previously acquired tendency to jump the barrier. Each subject was led into the experimental room on a leash and was coaxed into, or put into, one of the compartments. The door to the compartment was then locked and, since the gate was down (closed), the dog was thus completely confined. The lights above the compartments were on. The commutator motor was running, providing a low background noise. The dog was observed during a 10-min. pretrial "acclimation period."

Then, at the end of this period, the first pretest trial was conducted. The lights above the compartment in which the dog was confined were turned off by means of a foot pedal. Simultaneously, the experimenter raised the gate by pulling the counterweight on the pulley system. Therefore, the subject was in relative darkness but was, at the same time, exposed to the presence of the barrier, and immediately beyond it the other compartment was still illuminated. This situation will be referred to as the conditioned stimulus pattern (CS), or the "signal."

Each dog's reaction to the CS was observed. If, at the end of a 2-min. interval, the dog had not jumped out of the darkened compartment over the barrier into the illuminated compartment, the gate was lowered and the lights turned back on. Such a trial was recorded as "no response" with respect to jumping, but salient features of the dog's activity during his exposure to the CS were recorded in protocols.

One minute after the gate was closed, the second pretest trial was begun. It was identical to the first trial. Thus, the interval between presentations of the CS was 3 min. For most dogs, 10 pretest trials were conducted in order to make certain that each dog was not a "jumper" in this situation. (A few dogs were pretested for only five trials if they evidenced complete indifference to the CS.) After the tenth trial, the dog was left in the illuminated compartment with the gate down for a 10-min., postrial "cooling-off period." Then the dog was taken out of the apparatus and returned to his living cage. The 30 dogs used in this report were not "jumpers." That is, their "operant level" for jumping over a barrier approximately the height of their backs was close to zero.

Training. On the day after the pretest, each dog was subjected to avoidance training. The dog was brought into the experimental room in the same manner as for the previous pretest. He was, however, placed in the compartment opposite to the one in which he had been pretested. Just as had been done on the previous day, each dog was kept in one of the compartments, with lights on and gate closed, for a pretrial acclimation period of 10 min. duration. Then, at the end of this period the first training trial was begun. The CS was presented, just as has been described for the pretest trials. Ten seconds later, the grid on the floor of the darkened compartment was electrified with a voltage which was the *highest possible* without producing tetany of the dog's leg muscles. This shock voltage was monitored with the aid of the electronic voltmeter, and the current drawn at a *just-subtetanizing* stimulation level was approximately 10.0 to 12.5 ma. for most dogs. The voltage applied to the shock circuit was slightly different from dog to dog, since resistance conditions were not constant from dog to dog, and current flow necessary for tetany differed from dog to dog.

The shock was kept on in the darkened compartment until the dog jumped or scrambled over the barrier into the illuminated compartment. The gate was closed immediately following such an escape from shock, thus preventing any retracing on a given trial. If a dog failed to get out of the electrified compartment within a 2-min. interval following the presentation of the CS, the lights were switched on and the gate was lowered, accompanied simultaneously by termination of the shock. In such a case, the next trial was started after a 1-min. interval following the termination of the CS.

There was a standard CS-US temporal interval of 10 sec. maintained throughout all escape trials (the shock is the unconditioned stimulus). The time interval between trials (between successive presentations of the CS) was kept constant at 3 min. The latency of response was measured for each trial and was defined as the time interval elapsing between the presentation of the CS and the occurrence of jumping over the barrier. The latency was measured by stop watch to the nearest tenth of a second.

If a dog jumped within less than 10 sec. following the presentation of the CS, the gate was lowered to prevent retracing just as was the case in those trials on which the dog had a latency of over 10 sec. and was shocked. If a dog was moving toward the barrier at the end of the 10-sec. interval after the CS, the shock was withheld until the dog stopped moving toward the barrier. (This procedure was actually necessary on only 17 of 800 training trials.)

An *escape* trial or *shock* trial is defined as a trial in which the dog is shocked *before* he jumps the barrier. An *avoidance* trial is defined as a trial in which he jumps the barrier without being shocked. Thus, with the exception of the very few trials in which shock was withheld when the animal was approaching the barrier at the end of 10 sec., all avoidance trials have a response latency of 10 sec. or less, whereas all escape trials have a latency of more than 10 sec.

Ten training trials were conducted in each experimental session. Each session was concluded with a 10-min. posttrial cooling-off period. Then the animal was removed from the apparatus and returned to his home cage.

Certain aspects of the maintenance schedules for the animals were left uncontrolled. While the dogs were fed once a day, and usually had water in their home cages, it was true that sometimes they might be hungry and sometimes not hungry at the time of an experimental session. Usually the dogs were not thirsty. We believe that the variations in the maintenance schedule do not reflect themselves in our data on avoidance learning motivated by intense electric shock.

RESULTS

1. *Definition of Acquisition*

In *avoidance learning* it is very difficult to point out logically where the acquisition phase ends and extinction begins. This is so because the delivery of the unconditioned stimulus (shock) to the subject is contingent upon the subject's aversive movements. Thus, if one considers the *shock* to be the reinforcing agent necessary for acquisition, one might argue that extinction of avoidance begins when the aversive movements are such that shock is avoided. However, in terms of the experimenter's own operations, the acquisition-extinction distinction is not as easy to define as one might think. Suppose a subject suffers a *reversal;* that is, after avoiding successfully for a few trials, he has a long latency and is shocked before he responds with aversive movements. Where, then, would extinction begin? Also, if one takes an anxiety-reduction point of departure, the point at which acquisition ends and the extinction process begins is even more ambiguous. We cannot ascertain accurately at present whether or not anxiety reduction takes place after a given response on a given trial, so to define presence or absence of reinforcement in those terms is not feasible.

We have taken the stand that an arbitrary criterion is probably the best we can do—one based on how many times the experimenter has *not* shocked the dog. For all animals, if no shocks have been received for 10 consecutive trials, the acquisition phase is over; the animal has met our criterion for avoidance learning. And from then on, *no matter what the dog happens to do in the presence of the signal, he will never receive the shock*. The state of affairs that defines the extinction procedure is thus the *impossibility of receiving a shock* in either compartment, no matter how long the dog may delay jumping after presentation of the CS.

Despite such a definition of the onset of extinction, the *response characteristics* of our dog suggest that, in traumatic avoidance learning, rarely, if ever, do we enter a *true extinction* phase, a sequence of trials where no reinforcement occurs. For this reason, we will present behavioral data for several trials beyond the point at which the 10 avoidances out of 10 trials criterion has been met. In a later paper, we shall discuss extinction characteristics at greater length, but it should be clear here that we do not believe that there is any strongly tenable distinction available at present which would enable us to separate acquisition procedure from extinction procedure in the case of traumatic avoidance learning.

2. *Latencies of Jumping Responses*

a. Beginning of training as anchor point. The mean reciprocal of latency for the 30 dogs is plotted for each trial in Figure 2. Note that a logarithmic ordinate is used; this is merely for convenience. The median reciprocal of latency for the 30 dogs is plotted for each trial in Figure 3. Notice that as the response latencies decrease, the logarithms of the reciprocals of response latency increase in magnitude. The right-hand ordinate

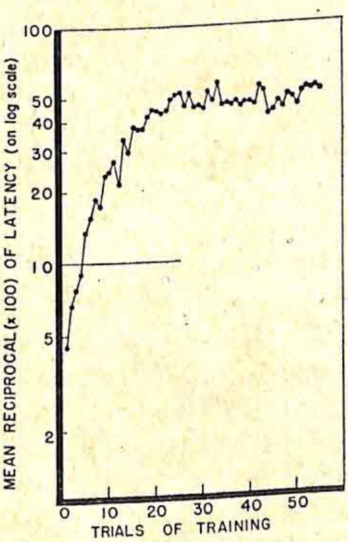


Fig. 2. Mean reciprocals of latency, trial by trial, for the 30 dogs.

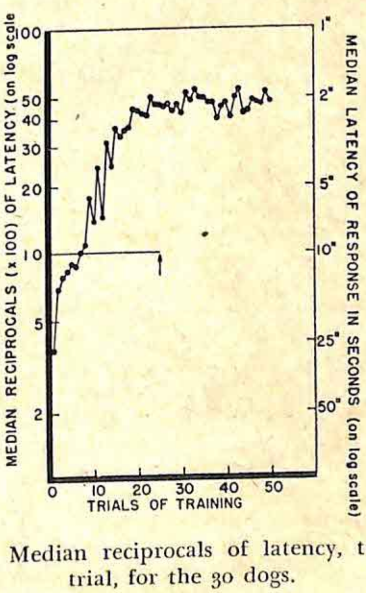


Fig. 3. Median reciprocals of latency, trial by trial, for the 30 dogs.

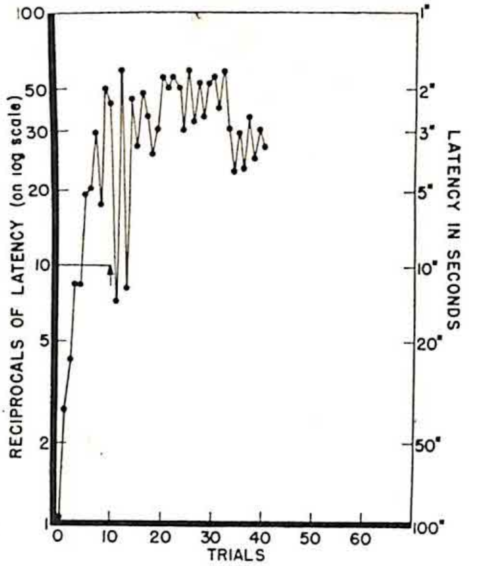


FIG. 4. An individual acquisition curve for a fairly typical subject. Note the two long latencies after the criterion has been met, followed by gradually shortening latencies.

translates the median reciprocals into the original latencies in seconds. For means, however, the right-hand ordinate is an approximation, since the antireciprocal of the mean reciprocals may not be the same as the mean raw latency. It may be convenient for the reader to interpret the graphs as indices of response strength, or response velocity. As response latencies decrease, response strength increases.

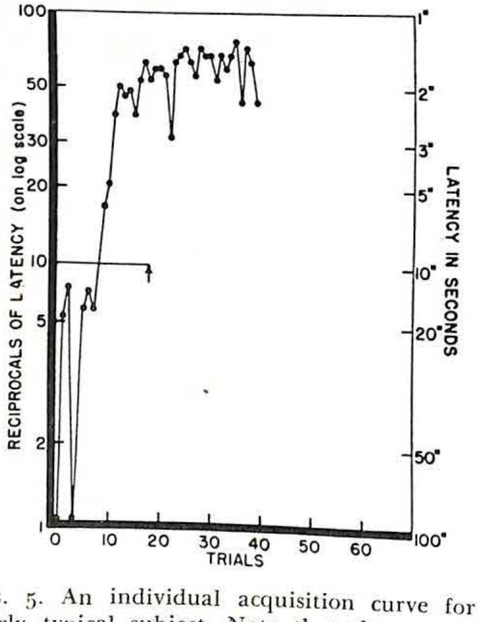


FIG. 5. An individual acquisition curve for a fairly typical subject. Note that the transition from escape responses to avoidance responses is final and complete, with no reversals. Here is an instance of avoidance learning which is not a case of irregular or aperiodic reinforcement.

We have found that the distribution of latencies on given trials in our experiment may be either bimodal, skewed, or normal, depending upon the phase of training. Therefore, ordinary techniques of data combination led to measures of central tendency which were definitely not representative of most of the individual animals in our experiment. In view of this we have selected a set of representative individual acquisition curves, Figures 4-9.

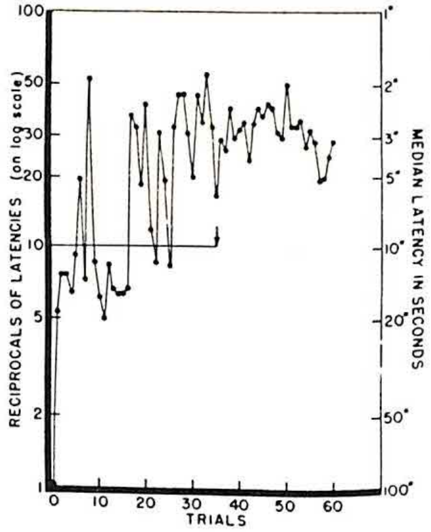


FIG. 6. An individual acquisition curve for a very slow learner. Note the large number of alternations from escape to avoidance responses, from avoidances to escapes.

In all of the graphs (Fig. 2-9), the horizontal line at 10 sec. (or at 10 on the scale of 1/latency × 100) indicates the boundary between escape and avoidance. Above the horizontal line, the points indicate a latency of response to the *con-*

ditioned stimulus alone, with no shock administered. Below the line, the points represent a latency of response to the US as well as to the CS. The distance that each point below the horizontal line lies from the horizontal line is the number of seconds of shock received while the animal was failing to get out of the "hot" compartment. These points represent escape or shock trials.

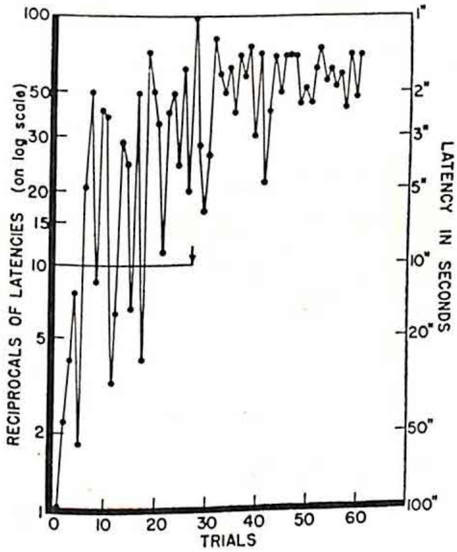


FIG. 7. An individual acquisition curve for a very gradual learner.

The vertical arrows in Figures 2-9 indicate the trials on which the criterion of acquisition was reached (10 avoidances in a row). After the arrow has appeared, the dogs are under extinction rules—they *cannot* receive shock for latencies of response to the CS of 10 sec. or more. They are now "safe," no matter what their responses are like.

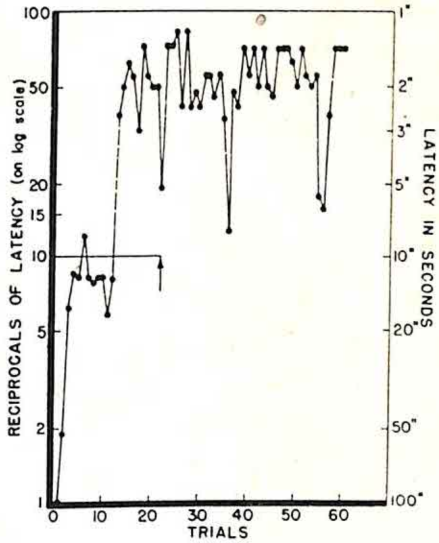


FIG. 8. An individual acquisition curve for a very sudden learner. Note the plateau of escape or shock trials, suddenly followed by avoidance responses of regular, short latencies.

An examination of the six individual records (Fig. 4-9) would lead one to suspect that the function in Figure 3 is *more representative of what actually occurred in our experiment* than is the function in Figure 2. That is, combination of data through use of *means* distorts our data more than with medians. This is very interesting on several counts. If we had generalized solely from the graphs of *individual* performance, we would have described a trend more simi-

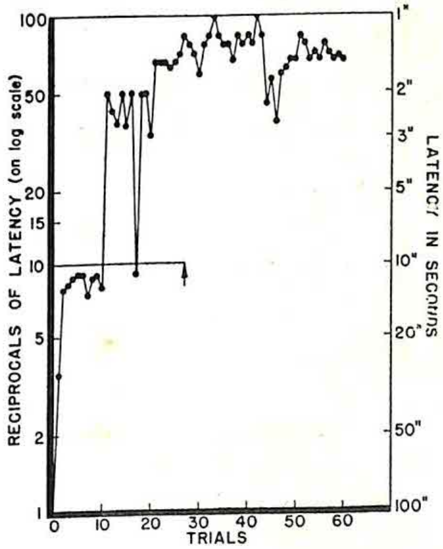


FIG. 9. An individual acquisition curve for a very sudden learner. Note the sudden transition to short-latency avoidance responses.

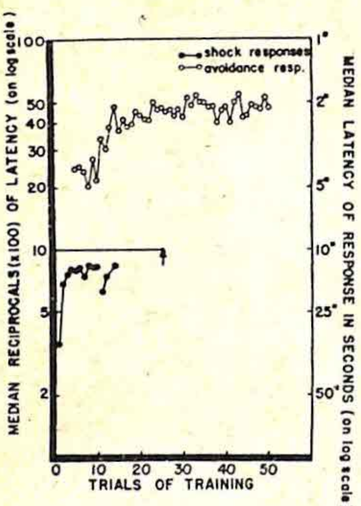


FIG. 10. Median reciprocals of latency for shock (escape) responses and avoidance responses considered separately. Note that the latency plateau for shock responses is reached after the first few trials, while the latency plateau for the avoidance responses is reached more gradually. Note also the absence of response latencies in the range between 5 and 10 seconds.

lar to that indicated by the *medians* of Figure 3. (This is what many physiologists typically do, without resort to statistical operations for combining data.)

An inspection of the individual acquisition curves reveals to some extent why the combined data, especially the means, are unrepresentative. First, the animals learn at different rates. In addition, there is a suggestion that we have a bimodal distribution of latencies because of a scarcity of latencies between 5 and 10 sec. Finally, we felt that there was an over-all tendency for the animals to learn abruptly, and if this were true, together with the differential learning rates, then such a tendency would be obscured by combining data in conventional ways. Many of the methods of analysis which we shall present have been designed to clarify the extent to which this tendency to learn abruptly exists. If there is a strong tendency for our animals to learn abruptly, then the following should be true: (*a*) There should be a sudden change from 100% escape responses to 100% avoidance responses; and (*b*) the latency of *response to the shock* on escape trials should be approximately equal to the latency of *response to the signal* on the first avoidance trials—that is, there should be no gradual shortening of latencies in making the transition from escape reactions to avoidance reactions.

Since the combination of latencies for escape and avoidance responses by different animals on any given trial will give intermediate values for the means, it seemed reasonable to separate escape and avoidance responses. Such a procedure is indicated for medians in Figure 10. Both of these graphs show that *escape* latencies change very little after the first few trials. On the other hand, the *avoidance* latencies do not begin in the vicinity of 10 sec. and gradually diminish; rather, avoidance responses begin in the vicinity of 5 sec. and decrease gradually to an asymptote.

In the first four trials, very few animals avoided. From Trial 5 to 7 there was a considerable increase in the proportion of the animals avoiding (43% on Trial 5, 37% on Trial 6, and 57% on Trial 7).

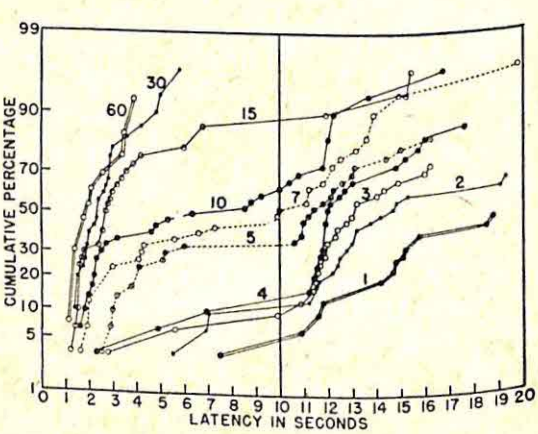


FIG. 11. Cumulative frequency of occurrence curves plotted on probability paper, showing the frequencies of occurrence of different response latencies at differing numbers of trials.

Hence, at this point in learning, if most of the animals are making an abrupt change from relatively long latencies to relatively short ones without a gradual diminution of latency, then the distribution of latencies on those trials should be bimodal.

A convenient way of demonstrating the characteristics of several different frequency distributions on one graph is the use of cumulative frequency distributions. If these are plotted on probability paper, they clearly demonstrate the degree of departure from a normal curve. (In such a graph a normal distribution will yield a linear function.) Figure 11 presents these curves for a selected sample of trials. For the first four trials, the frequency distributions of latencies are skewed, tending to pile up progressively at a latency of about 12 sec. On Trial 5, along with a marked increase in the number of animals avoiding the shock, a striking change in the shape of the frequency distribution takes place. There is a sudden increase in the percentage of latencies in the 2 to 5-sec. range. At the same time, there is *no increase* in the percentage of latencies in the 5 to 10-sec. range. Those latencies greater than 10 sec. continue in their tendency to pile up at about 12 sec. This bimodality gradually disappears as the number of training trials increases, so that by Trial 15 we again find a skewed distribution, piling up now in the 2 to 3-sec. range. This degree of skewness progressively diminishes, as illustrated in the plots of Trials 30 and 60.

A general, though not complete, description of acquisition of avoidance in this experiment can be derived from combining the implications of Figures 10 and 11. It would go something as follows: On the first four trials, most animals are getting considerable shock but their response latencies are decreasing. On Trial 5, a number of the animals shift over to respond quickly to the signal alone, while the rest continue to receive the shock. Finally, all of the animals avoid with very short latencies.

b. First avoidance trial as anchor point. Since the trial on which the first avoidance response occurs varies from animal to animal, combining the data on the basis of the number of trials of training obscures the characteristics of the first avoidance, as well as the responses preceding and following it. Therefore, we constructed a new baseline which is defined as *trials after the first avoidance,* and we combined the data of all animals for each of these trials.

When the escape latencies and avoidance latencies for *trials after the first avoidance* are unsegregated, as in Figure 12, an acquisition curve is obtained which would lead one to believe that there is a gradual transition from long to short latencies and that there is a high frequency of response latencies in the 5 to

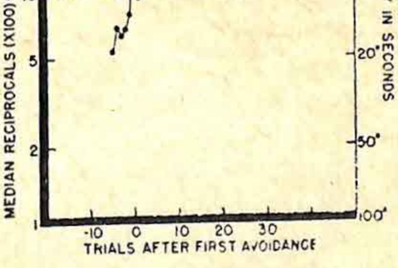


FIG. 12. Acquisition of avoidance as represented by the median reciprocals of response latency for trials after the first avoidance response. Escapes (shock trials) and avoidances are combined in determining the medians.

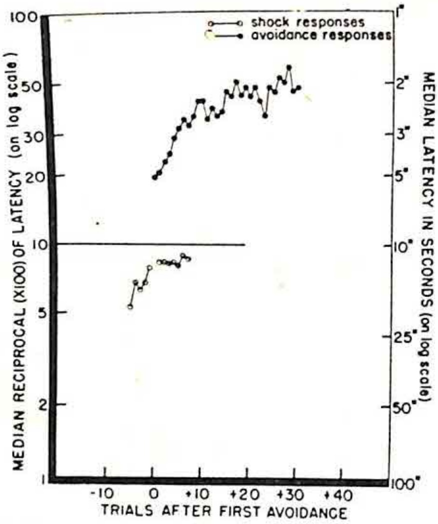


FIG. 13. Acquisition of avoidance as represented by median reciprocals of response latency for trials after the first avoidance response. Note that the shock trials and avoidance trials are treated separately.

10-sec. range. This impression, however, would be erroneous. When the escape or shock responses are treated separately from the avoidance or nonshock responses (Fig. 13), the median latency of the first avoidance response is approximately 5 sec. Thereafter, on successive trials following the first avoidance response, the latency of escape responses remains fairly stable at about 12 sec., but the latency of avoidance responses progressively diminishes. It is interesting to note that after an animal has made his first avoidance, he responds very quickly to shock. Immediately prior to the first avoidance the latency of the escape responses was still decreasing.

The cumulative frequency distributions for a selected sample of trials after the first avoidance are shown in Figure 14. The first avoidance trial (0) seems to be a transitional trial. On the trials before the first avoidance, the distributions are highly skewed, tending to pile up at a latency of about 12 sec. The trial after the first avoidance (+1), in contrast, shows a distinctly bimodal distribution of latencies. The distribution of latencies on the first avoidance trial is intermediate between the highly skewed type immediately before the first avoidance and the definitely bimodal type immediately after the first avoidance trial. Thus, with the exception of the first avoidance trial, the trials immediately following those on which 100% of the animals were failing to avoid exhibited bimodal distributions. On these trials after the first avoidance, if an animal failed to avoid, he jumped very quickly when shocked, and if he made an avoidance response, he jumped very quickly to the signal alone. There are very few responses in the 5 to 10-sec. range. Figure 14 shows this quite clearly.

c. Last shock trial as anchor point. Since the number of trials needed to meet the criterion for acquisition of avoidance learning (10 consecutive avoidance responses) differed from animal to animal, combining the data on the basis of the number of training trials obscures the characteristics of the trial on which the last shock was given, and those "criterion" trials directly following it. Therefore, we constructed another base-

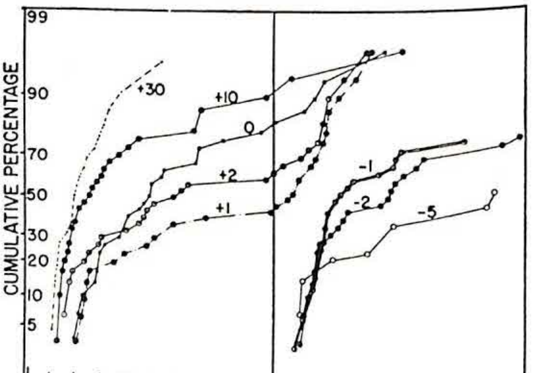


FIG. 14. Cumulative frequency distributions of latencies for various numbers of trials after the first avoidance response. The ordinate is on a probability scale.

line which is defined as *trials after the last shock*, and we combined the data for all animals for each of these trials.

When the latencies for all 30 animals for *trials after the last shock* are combined, leaving escape and avoidance responses unsegregated, the acquisition curve for median response latencies for trials after the last shock is quite striking. (See Fig. 15.) When the latencies for the shock and avoidance responses are separated, we obtain the curves plotted in Figure 16. The separated curves show clearly that the latencies of the avoidance trials were gradually decreasing *before* the trial on which the last shock was given. This indicates that, despite the relatively abrupt appearance of quite short avoidance latencies, these latencies diminish progressively with increased number of training trials. Interestingly enough, *the latencies of avoidance responses after the last shock has been received by all animals continue to decrease in an orderly manner,* approaching a median asymptote at approximately 1.8 sec. As was the case for the other baselines, the *shock response* latencies change little during those trials directly preceding the last shock trial.

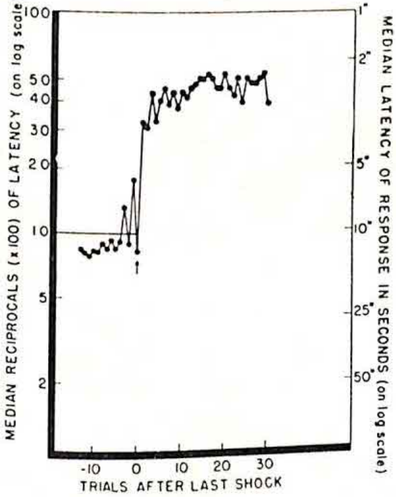


FIG. 15. Acquisition of avoidance as represented by median latencies for trials after the last shock (escape) trial. The shock and avoidance responses have been combined in determining the medians.

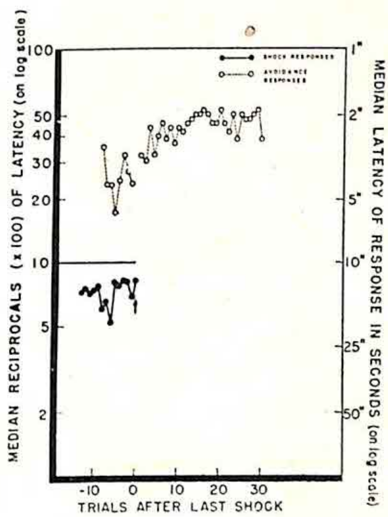


FIG. 16. Median reciprocals of latency for trials after the last shock when the shock and avoidance responses are separated.

The cumulative frequency curves for a selected sample of trials after the last shock are shown in Figure 17. The trials immediately preceding the trial before the last shock show the familiar bimodal distributions. The frequency distribution for the trial before the last shock (−1), however, seems to be transitional in

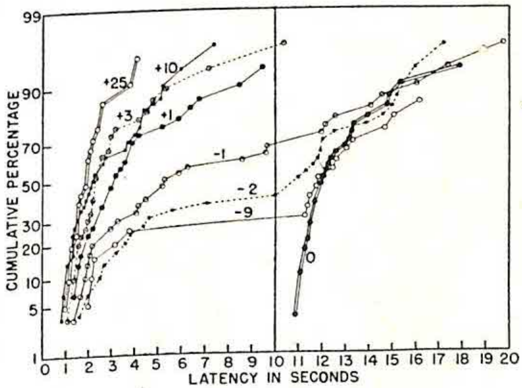


FIG. 17. Cumulative frequency distributions of latencies for different numbers of trials after the last shock trial.

shape, in that it is neither highly skewed nor bimodal in character. In contrast, the last shock trial (0) is definitely skewed with a piling up of response latencies at 12 sec. The trial after the last shock (+1) appears to have the same type of frequency distribution as the last shock trial (0), *but the median latency of the trial after the last shock is 9.1 sec. shorter than the median latency of the last shock trial.* The fact that the magnitude of the difference between these two medians approximates the magnitude of the CS-US interval used in training is probably not accidental, and the significance of this finding will be discussed later.

3. *Sequential Patterns of Escapes and Avoidances*

In the section above we have discussed response latencies. It should be pointed out that this has not been the accepted way of reporting the data of avoidance learning experiments (see, for example, Brogden, 2, 3). Usually the ordinate of the learning curve has been in terms of the percentage of avoidances in blocks of trials of arbitrary length, though there have been exceptions (see, for example, Miller, 6, 7). In our experiment, learning took place so rapidly that this method of analyzing the data was not feasible. Neither was the Vincent curve technique feasible. Instead, it seemed more reasonable to supplement the latency data with an analysis of sequential patterns of escape responses and avoidance responses, *without regard to latency magnitudes within each class of responses.*

Before examining the specific sequential patterns, an over-all summary of occurrences of escapes and avoidances for

TABLE 1

PER CENT OF ANIMALS AVOIDING ON EACH TRIAL, FOR ALL THREE BASELINES

Trials of Training			Trials after the First Avoidance			Trials after the Last Shock		
Trial No.	*N*	%	Trial No.	*N*	%	Trial No.	*N*	%
1	29	0.0	−1	30	0.0	−13	13	23.1
2	30	10.0	0	30	100.0	−12	14	28.6
3	30	10.0	+1	30	50.0	−11	17	18.8
4	30	13.3	+2	30	66.7	−10	19	21.0
5	30	43.3	+3	30	56.7	−9	20	25.0
6	30	36.7	+4	30	73.3	−8	23	47.8
7	30	56.7	+5	30	63.3	−7	26	34.6
8	30	60.0	+6	30	66.7	−6	28	50.0
9	30	66.7	+7	30	73.3	−5	30	43.3
10	30	76.7	+8	30	80.0	−4	30	56.7
11	30	73.3	+9	30	83.3	−3	30	56.7
12	30	60.0	+10	30	90.0	−2	30	40.0
13	30	90.0	+11	30	93.3	−1	30	66.7
14	30	70.0	+12	30	93.3	0	30	100.0
15	30	86.7	+13	30	96.7	+1	30	100.0
16	30	90.0	+14	30	96.7	+30	14	100.0
17	30	90.0	+15	30	96.7			
18	30	93.3	+16	30	96.7			
19	30	96.7	+17	30	100.0			
20	30	100.0	+18	30	100.0			
25	30	96.7	+19	30	96.7			
30	30	100.0	+20	30	100.0			
35	21	100.0	+30	22	100.0			
40	21	100.0						
45	14	100.0						
50	14	100.0						

TABLE 2

RELATIONSHIPS AMONG ESCAPE AND AVOIDANCE RESPONSES

Relationship	Mean	Median	Range
Number of escapes before first avoidance	4.5	4.0	1–8
Number of escapes before second avoidance	6.5	6.0	2–14
Number of trials of training to and including the last shock	12.4	12.0	6–25
Total number of shocks received	7.8	7.0	4–13
Number of trials after first avoidance up to and including the last shock	6.8	6.5	0–19
Number of shocks received after first avoidance response	3.3	3.0	0–9
Number of alternations of escapes and avoidances before meeting criterion of acquisition	5.5	5.0	1–13

the three types of baseline will serve as orientation to the data. An examination of Table 1 will reveal the general characteristics of the acquisition of avoidance without regard to response latency.

The sequential patterns of escape and avoidance responses stand out more clearly if we measure the central tendencies for escapes and avoidances to occur in certain sequential positions, regardless of response latency characteristics. These central tendencies are shown in Table 2.

If we take the median sequential relationship as representative of the "typical" animal in our experiment, we arrive at the following sequential pattern: four successive escapes (shocks) on the first four trials of training; the first avoidance response on Trial 5; two escape and four avoidance responses distributed over Trials 6 through 11, with two alternations of escapes and avoidance within this sequence; the last shock appears on Trial 12, followed by 100% avoidance responses. This sequence pattern makes no reference to magnitudes of response latency within the two classes of response.

4. *Amount of Shock Received*

Since the amount of shock received by a given animal (in seconds of shock) varied from animal to animal, depending upon many factors, it seems important to assess its relevance to the behavioral aspects of individual animals. The total number of seconds of shock varied from 10.1 sec. to 454.9 sec., with a mean of 93.5 sec. and a median of 56.8 sec. The first quartile of this distribution was 24.5 sec., and the third quartile was 146.3 sec. That the total seconds of shock received were related to sequential aspects of the course of acquisition is suggested by the fact that the total seconds of shock were negatively correlated with the number of alternations of escapes and avoidances. The *more* shock an animal receives, the *less* tendency there is for him to be sequentially erratic. The correlations coefficient (rho) is −0.44.

Most of the relationship between amount of shock received and number of alternations is due to shock received *on those trials preceding the first avoidance* response. The correlation of seconds of shock before the first avoidance with number of alternations is −0.38. The correlation between the number of shocks before the first avoidance and number of alternations is −0.42. While these correlations are low, they suggest that *relatively large amounts of shock before the first avoidance tend to decrease sequential variation during the trials following the first avoidance.* Other correlations relating seconds of shock and number of shocks before the first avoidance to other aspects of acquisition were low, but they were consistent with the implications of the correlations cited above.

5. *Behavioral Observations*

In the protocols, we tried to record the following kinds of observations: (*a*) gross skeletal movements and vocalization; (*b*) position within the apparatus at selected points in time; (*c*) signs of autonomic reactions, such as: salivation, defecation, urination, shivering, panting, pupillary changes, and respiratory rate; (*d*) anticipatory reactions, such as position shifts directly following presentation of the CS, and autonomic signs, vocalization, and skeletal movements directly following the CS. Clearly, such observations give us little information about the events taking place in the nervous system or details about the animal's motivational state. However, it is possible that such observations may be related to the occurrence or nonoccurrence of avoidance responses. In the present study, we are especially interested in the possible relationship between these behavioral observations and the occurrence of the first avoidance response and the last shock response.

The animals' responses to the unconditioned stimulus (subtetanizing shock) were quite uniform. (A description of these responses was given in the introduction.) However, it was still possible for the experimenters to rank the animals in order of intensity of reaction to the shock. This ranking procedure was carried out in an arbitrary fashion by assigning weights to the occurrence of specific emotional signs. The signs which were recorded in the protocols, and the weights which were assigned to each, are listed in Table 3. Using this scale, an emotionality score could be computed for each dog, and this score could be then correlated with other observations and measures derived from the data.

TABLE 3

CLASSIFICATION OF EMOTIONAL SYMPTOMS AND WEIGHTS USED IN COMPUTING THE EMOTIONAL INDEX

Emotional Behavior	Weight
Urination	3
Defecation, breaking wind	3
Salivation, drooling	2
Yelping, shrieking	2
Attacking apparatus, clawing at it	3
Shivering, shaking, trembling	3
Pupillary dilatation	3
Whining	2
Posturing	2
Restlessness, agitation	1
Barking	1
Panting	1
Scrambling on grid bars	1

The results of the analysis of emotional reactions were highly instructive. The mean trial on which the first anticipatory emotional signs occurred was 3.6. It will be remembered that the trial on which the first avoidance response occurred, as listed in Table 2, was 5.5 (a mean value of 4.5 shock trials before the occurrence of the first avoidance response). Thus, on the average, the dogs tended to exhibit some type of overt emotional reaction to the CS on a trial approximately two trials before the trial on which the first instrumental avoidance response to the CS occurred. It should be emphasized that the analysis showed that the anticipatory emotional signs were predominantly skeletal rather than visceral in nature: symptoms such as restlessness and agitation, whining and barking, and, in several dogs, shying away from the gate and retreating to the rear of the compartment. There is great likelihood that many anticipatory visceral reactions were unobservable without instrumentation. Since the raising of the gate signalled the oncoming shock, it is not surprising that some animals retreated from it; in a few cases, primitive "perceptual defense" was observed when a dog would shy away from the gate and hide his head in a corner of the compartment so that the visual stimuli from the open gate were effectively shut out.

Anticipatory signs in the presence of the CS tended to be more obvious in the fast learners who received relatively few shocks. The correlation coefficient for the relationship between the number of shocks before the first avoidance response and the emotionality score was −0.39. This might indicate that the more upset a dog appears to be (in the judgment of the experimenters) during the first few trials of training, the quicker will the first avoidance response appear. This could be interpreted to mean that there is a tendency for the more frightened dogs to learn more quickly. However, the correlation between the emotionality score and seconds of shock received before the first avoidance was only −0.16. While the difference between the rho of −0.39 and −0.16 was not significant, it is possible that the *frequency* of shocks is not the same as *seconds* of shock received with regard to psychological significance.

In general, analysis of the emotionality scores both before and after the dogs reached the criterion for learning (ten avoidance responses in 10 trials) reveals that the dogs were more upset before the criterion was met than they were afterward. But it should be emphasized that several of the dogs remained in an extremely agitated state after they were regularly avoiding the shock.

The protocols contained a brief description of the behavior of the animals during the 3-min. interval between trials. There was a strong tendency for stereotyping to develop. For example, immediately after jumping a dog might position himself in a specific part of the apparatus, facing his body and head in a fixed direction, and he might maintain this position until the next presentation of the CS. In opposite compartments of the apparatus, such stereotyped behavior would often be symmetrical or a mirror-image. When such a degree of stereotyping was observed, it was usual to find that many previously exhibited emotional signs were no longer evident.

DISCUSSION

We believe that the data of this experiment are most efficiently discussed within the framework of a two-process learning theory. The theoretical writings of Skinner (15, 16), Schlosberg (13), and Mowrer (8, 9) are prominent in the development of our own point of view. We assume that in the establishment of avoidance learning two processes are operative: (*a*) the process of classical conditioning governed by the principle of stimulus contiguity, and (*b*) the process of trial-and-error learning governed by the S-R reinforcement principles.

The classical conditioning component is responsible for the appearance of *emotional* responses elicited by the conditioned stimulus. Some of the attributes of the original unconditioned response to the noxious stimulus come under the control of the once-neutral CS. The animal then appears to be emotionally upset or anxious, exhibiting a wide variety of reactions that include both skeletal and visceral components. We do not follow Mowrer's relegation of autonomic reactions to classical conditioning and skeletal responses to trial-and-error learning. Rather, we adhere more closely to Schlosberg's conception that diffuse conditioned emotional responses may develop under conditions of stimulus contiguity. These classically conditioned responses give rise to a stimulus pattern having drive properties. It is this stimulus pattern to which Mowrer (8) and Miller (6) refer in describing anxiety as an acquired drive.

The reinforcement of learned instrumental avoidance responses comes about through drive reduction. Early in the learning process when the animal is escaping from shock, the instrumental act removes the US, as well as the CS. Drive reduction then consists of reduction in the intensity of both pain and emotional upset. Later, when the animal is avoiding the shock, drive reduction consists of reducing the intensity of the emotional upset by removing the CS.

Our formulation of a two-process theory enables one to explain without further assumptions: (*a*) the development of conditional emotional responses during the course of avoidance training, and (*b*) the maintenance of avoidance responses once they have occurred. However, the mechanism whereby the animal makes the *transition* from escape to avoidance behavior is usually neglected in theoretical accounts of avoidance learning.

With repetition of the CS-US sequence (during the escape phase), the emotional reaction to the

CS increases in amplitude or intensity. Thus, the response pattern to the CS will gradually approximate more closely the emotional response to the US, and the response-produced drive stimulus pattern which follows the CS will gradually approximate more closely the response-produced drive stimulus pattern which follows the US. During the escape phase the instrumental act has consistently taken place in the presence of a stimulus pattern composed of the CS, US, and emotional upset. When the CS will elicit the same sort of emotional upset, the instrumental act will tend to take place even though the US has not occurred. Presumably the stimulus generalization principle will account for this phenomenon.

The animal which previously responded instrumentally to CS, US, and upset now responds instrumentally to CS and upset. Thus a two-process theory will explain the occurrence of a transition from escape to avoidance responses.

However, whether the transition will be abrupt or gradual is not derivable from such a theory. Nor, indeed, are such predictions possible on the basis of cognitive theory, S-R reinforcement theory, or S-R contiguity theory. The degree of abruptness of avoidance learning can be defined in several ways, but all such definitions seem to fall into two classes: (*a*) those that characterize the rate of change of latency of the instrumental responses, and (*b*) those that characterize the rate of change from 100% escape responses to 100% avoidance (sequential pattern).

It is interesting that a two-process theory allows a prediction about the magnitude of latency of the first avoidance response while not making a direct prediction about abruptness of latency change. Other theories do not seem to lend themselves to making even so limited a prediction. If, as we argued above, the first avoidance response appears as a reaction to the stimulus pattern which accompanies the emotional upset, then the latency of the instrumental response will be the summation of two quantities. The first quantity will be the time between the presentation of the CS and the appearance of the conditioned emotional response. This latency quantity will presumably be inversely correlated with the level of intensity of the conditioned emotional response (1, 12, 20). With a fairly high level of emotional intensity we can reasonably expect a latency of about 1.5-2.5 sec. The second latency quantity is the characteristic time required for performing the instrumental act. This can be estimated from the latency of escape responses. In our data the median latency for the trial before the first avoidance was 12.6 sec., indicating that the instrumental response required 2.6 sec. following the onset of the shock. Adding the two quantities together gives an expected range from 4.1 to 5.1 sec. for the total latency of the first avoidance response. The actual median latency of the first avoidance response was 5.1 sec.

It is interesting to note that whether or not this transition from the escape latency to the avoidance latency is judged to be abrupt depends on the magnitude of the CS-US interval. If the magnitude of the latency of the first avoidance is predetermined by the two processes described above, then with a long CS-US interval the rate of change of latency will seem great on the first avoidance trial. If, however, the CS-US interval is relatively short (for example, if it had been 5 sec. in our experiment), the transition from escape to avoidance will seem to be characterized by a small change in latency. These theoretical considerations point up the importance of carrying out experiments systematically varying the CS-US interval in avoidance learning.

It is to be expected that fluctuations in the physiological state of the animal will affect the intensity of the conditioned emotional response and stimulus pattern. Such changes in intensity will be reflected in fluctuations in the latency of the instrumental response. Then, too, owing to changes in orientation of the animal in the apparatus, the degree of effective exposure to the CS will vary somewhat from trial to trial. The time mechanically necessary for the animal to get over the barrier will also vary with orientation in the apparatus. These fluctuations associated with orientation will be greater in the early stages of training before relatively stereotyped behavior in the interval between trials has developed.

If there is fluctuation in latency, this may produce reversals from avoidance to escape. Such reversals will lead to additional reinforcement of the instrumental act due to reduction of anxiety and pain. Therefore, when the next avoidance response does occur, its latency should be shorter than that of the first avoidance response. Also, the animal may show a lowering of latency associated with adjustments in orientation in the apparatus and increased proficiency in making the instrumental response. In our experiment, the median latency for the second avoidance trial was 4.1 sec., a full second shorter than that for the first avoidance trial.

With repeated reinforcement due to anxiety reduction on avoidance trials, we would expect that the latency of the instrumental act should decrease gradually even after no more shocks are being administered. Our data on trials after the last shock bear out this expectation. The median latency for the first five trials after the last shock was 3.1 sec. The median latency for Trials 26-30 after the last shock was 2.1 sec. It is interesting to note that on the first five trials after the first avoidance response, the median latency for escape responses, when they did occur, was 12.1 sec. Thus, the instrumental response component, at a comparable level of training, is exactly the same (2.1 sec.) no matter whether the response is elicited by the CS or by the US. When the instrumental response is elicited *at this time* by the CS, it can no longer be argued that a conditioned emotional reaction is operative in eliciting the instrumental response. This can be deduced because the instrumental act occurs before, or simultaneously with, any emotional reactions elicited by the CS. When this occurs, it appears that the animal is acting cognitively.

The reversals from avoidance to escape, or from escape to avoidance appear to be an all-or-none phenomenon, as though the control of the instrumental component were *switching* back and forth from the CS to the US. The median shift of latency from the last shock trial to the first trial after the last shock (first criterion trial, +1) was 9.1 sec., and an all-or-none conception would require this difference to be the magnitude of the CS-US interval, or 10.0 sec. The difference between 10 sec. and 9.1 sec. in this case is, however, not statistically significant. Here again, the carrying out of an experiment which systematically varies the CS-US interval from group to group appears to have strategic importance. For example, using the experimental procedures which produced our data, if the CS-US interval were 20 sec. instead of 10 sec., then the shift in latency from the last shock trial to the subsequent trial should be 20 sec. The first avoidance trial should exhibit a latency of about 5 sec., decreasing gradually to a latency of about 2 sec. at 30 trials following the last shock trial.

The theoretical significance of the "switching phenomenon" is hard to assess. Most cases of classical conditioning (see 1, 4, 5, 20) which seem to fit a stimulus-substitution paradigm are characterized by *gradual* growth in the magnitude of the CR. If the switching phenomenon is a special case of a stimulus substitution process, then under what conditions does it occur? Cognitive theorists might interpret the switching phenomenon as an example of insight. But, then, why do we see reversion to escape reactions after the animal has demonstrated his capacity to avoid the shock? An S-R reinforcement or S-R contiguity interpretation might emphasize the role of competing responses in accounting for the data; such theories might argue that the experimental situation, and especially the US, tends to elicit responses which may be incompatible with the instrumental avoidance response. Such an interpretation would indeed account for plateaus in the learning process, but would not lead to a deduction of the switching phenomenon. As a matter of fact, response incompatibility might suggest the occurrence of a plateau at intermediate latencies in our experiment. Such was not the case in our data. In short, we do not feel that a fully adequate theoretical explanation of such findings is available at present.

An examination of the sequential pattern of escape and avoidance responses reveals that three animals switched from 100% escape responses to 100% avoidance responses without a reversal. Seven animals had only one reversal and eight animals had only two reversals. Sheffield and Temmer (14) have argued that random or aperiodic reinforcement schedules are characteristic of avoidance training. They have explained the relatively slow course of extinction of avoidance responses in terms of the random pattern. However, our data indicate that such a pattern is *not* necessary in the normal course of avoidance learning; and, indeed, a fairly exact analysis of the actual sequential patterns is necessary in interpreting avoidance data. The fact that some animals *can* learn to *avoid* with the same schedule of reinforcement with the US as that used in *escape* training needs to be emphasized in theories which try to account for the difference between escape and avoidance learning.

The descriptive protocols produced some interesting relationships. Despite the incompleteness of the emotionality data, certain strong tendencies stood out, and they pose some special problems for theories of avoidance learning. It will be remembered that, on the average, the first overt emotional reaction to the CS occurred approximately two trials before the appearance of the first avoidance response. Mowrer (8, 10) has reported a similar finding, and both his observations and ours are compatible with a two-process theory. This relationship between the occurrence of conditioned emotional responses and the appearance of instrumental behavior is worthy of more study. While a two-process theory accounts for the development of classically conditioned emotional responses, and for the development of instrumental acts reinforced by drive reduction, the interrelationships between the two processes are not worked out. It is perfectly conceivable that there is a great deal of independence between the two processes under some conditions. An animal might become upset at a signal which had been associated with shock, but because the instrumental response which might eliminate the signal may be low in the response hierarchy, the animal may shift from ritual act to ritual act without developing any strong instrumental behavior. Furthermore, the response hierarchy can vary with differing external environmental conditions as well as differing internal emotional or drive conditions. Internal emotional conditions which are not directly arousable by the CS itself, but which are long-lasting and a function of the total situation, can determine the response hierarchy so that the act which might eliminate the signal may have either a low *or* high probability of being emitted. In avoidance conditioning, we have argued, there is a close interrelationship between the two processes, a relationship mediated by the assumption that conditioned emotional reactions have drive properties. But the determinants of the extent of this relationship have not been studied systematically. Our data reveal that the observer's prediction of the occurrence of the first avoidance is aided by the data on anticipatory emotional responses recorded in the protocols. The observer is also aided in his predictions by knowledge about the instrumental response as measured by the latency of escape; the shorter the escape response latencies, the greater is the likelihood of a shift to avoidance.

Another indication that the occurrence of emotional reactions is related to the learning of avoidance is the relationship between the emotionality scores and the sequential pattern. The correlation of -0.39 between the emotionality index and the number of shocks before the first avoidance suggests that, other things being equal, with greater emotional upset the animals make the instrumental response of avoidance sooner. However, there is no relationship (rho $= -0.06$) between emotionality and the total number of trials to meet the learning criterion of ten successive avoidances. Explaining such a finding is extremely difficult because several partial relationships seem to be operating in opposite directions. For example, while the most upset animals require the fewest shocks before the first avoidance, those animals receiving the fewest shocks in early trials tend to alternate between escape and avoidance. Thus, the most upset animals, receiving the least shock in early trials, should alternate more after the first avoidance. That the trend is in this direction is indicated by a rho of $+0.27$ between the emotionality index and the number of alternations between escape and avoidance.

Although a number of very general theoretical problems have been introduced to this discussion, we wish to emphasize that the particular conditions under which our experiment was carried out dictated, of necessity, the direction that the discussion took. It should be remembered that our animals were exposed to the following important conditions: (*a*) a highly traumatic unconditioned stimulus, (*b*) a CS-US interval of 10 sec., (*c*) a relatively difficult instrumental act in jumping over a high barrier, (*d*) a shuttle sequence which required the animals to return to the same compartment on alternate trials, (*e*) a CS which remained present until the instrumental act took place, and (*f*) the subjects were mongrel dogs. The results were a function of the experimental procedures, and the theory which seemed to fit our results best may prove to be inadequate under other conditions. Many variations of the conditions above will be involved in subsequent tests of the generality of both our findings and the theory we used to account for them. The procedures we did use produced behavioral data which posed challenging problems for theories of learning, and many new experiments were suggested in the course of considering these problems.

SUMMARY

An experiment in traumatic avoidance learning is reported in which dogs were trained to avoid a just-subtetanizing shock by responding to a signal which preceded the shock by a period of 10 seconds. A shuttle-box jumping response was reinforced as the instrumental avoidance reaction. The dogs received ten trials per day with a 3-min. interval between trials.

The results can be stated briefly: (1) During the *escape* phase of learning, when the animals were receiving shocks because they were not "anticipating," the latency of the jumping response decreased rapidly and did not change significantly after the third trial. (2) The first *avoidance* response, a response to the signal having a latency of less than 10 seconds, appeared (on the average) on the fifth trial; it was preceded at about the third trial by the appearance of conditioned emotional reaction to the signal. (3) The latency of the first avoidance response represented an abrupt decrease from the escape reaction latencies preceding it. (4) After the first avoidance trial the latency of avoidance responses continued to decrease in an orderly manner, while the latencies of the escape reactions did not change. (5) The latency of avoidance reactions continued to decrease after the trial on which the last shock was received. (6) The difference in latency between the last shock trial and the avoidance trial following it was 9.1 seconds, approximately the magnitude of the CS-US interval. (7) The sequential patterns of escape and avoidance reactions showed that some dogs make a complete change from 100% escapes to 100% avoidances without reversal. (8) Relatively large amounts of shock received during the first few trials tended to decrease the number of reversals from avoidance to escape on the trials following the first avoidance response. (9) Stereotyping of behavior was noted in conjunction with decrease in emotional reactions.

The data of this experiment were analyzed, with the aid of several new techniques for representing the course of change of latencies and the characteristic sequential patterns of escape and avoidance responses. These techniques helped to some extent in presenting data on abrupt changes in behavior.

The results were discussed within the framework of a two-process theory of learning. Several inadequacies in current learning theories were revealed in trying to explain our findings.

REFERENCES

1. Bitterman, M. E., Reed, P., & Krauskopf, J. The effect of the duration of the unconditioned stimulus upon conditioning and extinction. *Amer. J. Psychol.,* 1952, **65,** 256-262.
2. Brogden, W. J. Acquisition and extinction of a conditioned avoidance response in dogs. *J. comp. physiol. Psychol.,* 1949, **42,** 296-302.
3. Brogden, W. J., Lipman, E. A., & Culler, E. The role of incentive in conditioning and extinction. *Amer. J. Psychol.,* 1938, **51,** 109-117.
4. Hovland, C. I. The generalization of conditioned responses. IV. The effects of varying amounts of reinforcement upon the degree of generalization of conditioned responses. *J. exp. Psychol.,* 1937, **21,** 261-276.
5. Kellogg, W. N. Evidence for both stimulus-substitution and original anticipatory responses in the conditioning of dogs. *J. exp. Psychol.,* 1938, **22,** 186-192.
6. Miller, N. E. Studies of fear as an acquirable drive: I. Fear as motivation and fear-reduction as reinforcement in the learning of new responses. *J. exp. Psychol.,* 1948, **38,** 89-101.
7. Miller, N. E. Learnable drives and rewards. In S. S. Stevens (Ed.), *Handbook of experimental psychology.* New York: Wiley, 1951. Pp. 435-472.
8. Mowrer, O. H. On the dual nature of learning—a reinterpretation of "conditioning" and "problem-solving." *Harvard educ. Rev.,* 1947, Spring, 102-148.
9. Mowrer, O. H. Two-factor learning theory: summary and comment. *Psychol. Rev.,* 1951, **58,** 350-354.
10. Mowrer, O. H., & Lamoreaux, R. R. Fear as an intervening variable in avoidance conditioning. *J. comp. Psychol.,* 1946, **39,** 29-50.
11. Mowrer, O. H., & Miller, N. E. A multipurpose learning-demonstration apparatus. *J. exp. Psychol.,* 1942, **31,,** 163-170.
12. Pavlov, I. P. *Conditioned reflexes: an investigation of the physiological activity of the cerebral cortex.* (Trans. by G. V. Anrep.) London: Oxford Univer. Press, 1928.
13. Schlosberg, H. The relationship between success and the laws of conditioning. *Psychol. Rev.,* 1937, **44,** 379-394.
14. Sheffield, F. D., & Temmer, H. W. Relative resistance to extinction of escape training and avoidance training. *J. exp. Psychol.,* 1950, **40,** 287-298.
15. Skinner, B. F. Two types of conditioned reflex and a pseudo-type. *J. gen. Psychol.,* 1935, **12,** 66-77.
16. Skinner, B. F. *The behavior of organisms.* New York: Appleton-Century, 1938.
17. Skinner, B. F., & Campbell, S. L. An automatic shocking-grid apparatus for continuous use. *J. comp. physiol. Psychol.,* 1947, **40,** 305.
18. Solomon, R. L., Kamin, L. J., & Wynne, L. C. Traumatic avoidance learning: the outcomes of several extinction procedures with dogs. *J. abnorm. soc. Psychol.,* 1953, **48,** 291-302.
19. Solomon, R. L., & Wynne, L. C. Avoidance conditioning in normal dogs and in dogs deprived of normal automatic functioning. *Amer. Psychologist,* 1950, **5,** 264. (Abstract)
20. White, C. T., & Schlosberg, H. Degree of conditioning of the GSR as a function of the period of delay. *J. exp. Psychol.,* 1952, **43,** 357-362.

(Accepted for publication November 28, 1952.)

Vol. 67, No. 5 Whole No. 355, 1953

Psychological Monographs: General and Applied

Errors in Time-Study Judgments of Industrial Work Pace[1]

Kalman A. Lifson
Occupational Research Center, Purdue University

I. INTRODUCTION

Time studies are made to determine how much time should be required to accomplish specified elements of work. One of the steps in a time study is the calculation of the "normal time"—the time which the work element requires when performed at a "normal" pace. In determining the normal time, the time-study man records the amount of time actually used by the operator he is observing, and he also makes a judgment of the work pace of the operator. The actual observed time is adjusted in accord with the judgment of work pace so that the adjusted time is representative of a "normal" performance.

Usually the judgment of work pace is expressed as a percentage of "normal" work pace, and the adjustment is the multiplication of the observed time by this percentage. If, for example, the time-study man believes the observed worker is maintaining a pace that is 30 per cent faster than normal, he rates the performance for the observed work element at 130 (omitting the decimal) and multiplies the observed work time by 1.3.

The judgment is called by various names such as "leveling," "effort rating," "performance rating," "pace rating," or just plain "rating." In this paper the term "rating" will be employed. This rating may be defined as the process of comparing an observed work pace to some concept of a "normal" pace. Time studies are only as reliable and accurate as the ratings on which they are based.

Purpose of Investigation

This study examined the nature of the errors involved in ratings of work pace from several aspects: (*a*) by determining the magnitude and assignable causes of errors in pace ratings; (*b*) by determining the variation among the consistencies of ratings from different sources; and (*c*) by comparing the relative concepts of "normal pace" of workers with the concepts of the time-study men.

Sources of error. Factors which are the assignable sources of error in pace ratings are evidenced by constant differences in the average level of those ratings which are common to the factor. Specifically: (*a*) *different observers* may have different general concepts of normal work pace, and tend to rate either consistently high or consistently low; (*b*) *different jobs* may appear easier to some observers than to others, and this variation may be superimposed on the general high- or low-rating tendency of the observer; (*c*) *different workers,* maintaining the same tempo of elements per minute on a job, may also be rated differently; and (*d*) the normal times derived may differ when *different levels of pace* are observed—lower paces may be rated too high, and higher paces may be rated too low.

---

[1] This paper is based upon a thesis (5) submitted to the Graduate School of Purdue University in partial fulfillment of the requirements for the degree of Doctor of Philosophy, August, 1951. The research was done under the committee of Professors E. J. McCormick, chairman, Joseph Tiffin, N. C. Kephart, W. V. Owen, and M. E. Mundel.

order of presentation of the three reels of the film was reversed. A rest pause of about 10 minutes was taken by the raters between reels.

Method of Obtaining Worker Judgments

In order to be able to compare the time-study men's ratings of work paces with the workers' judgments of these paces, it was necessary to quantify the workers' judgments. These judgments represented the workers' relative evaluations of the paces on the four jobs. The quantification was done by having the worker compare, in dollars and cents, his feelings about the four jobs. The worker performed job 2 (stamping) at the normal pace, and was told that by working at that rate he could earn $1.00 an hour. Then he was given another job to perform at one of the five paces. He was asked, "If stamping like that you earned $1.00 an hour, how much should you be earning now?" After answering, he went back to the normal stamping, and then to another job pace to make the next comparison. This process was repeated until he had compared each job at each pace with the standard of stamping at normal pace.

Next, a different job, at a pace that the worker had equated to the normal stamping, by answering "$1.00 an hour" to the question of earnings, was chosen as the standard. The judgment process was repeated, this time using the new standard for comparison.

Finally, a third standard pace, on a third job, was chosen and the process was run again.

Thus, for each pace on each job, the worker made three judgments. The workers were told nothing about the relationship of the paces, or that they were being given the same paces over again. The job paces were presented in a random order.

Statistical Techniques

Analyses of variance (2, pp. 176-190) were made for the 1,200 time-study pace ratings obtained and for the 300 worker judgments. The interpretation of confidence levels was conservative because of the nonhomogeneous variances present. An analysis of covariance (2, pp. 333-355) was made between worker judgments and time-study pace ratings on each pace of each job.

Reliability of a category of ratings, and of worker judgments, was calculated as the standard deviation of the errors of the ratings in that category, and is expressed as the standard error of the ratings. Significances of differences among standard errors were tested by Bartlett's test (2, pp. 195-197).

III. RESULTS

Reasonably conclusive answers to the questions proposed in the introduction to this study are provided by the data. The interpretation of this summary of results is modified by the following limitations of the study: the time-study men, although they professed the same verbal concept of normal pace, were not from the same plant; in addition, the time-study men may have experienced fatigue from so many ratings; further, the rating on any one shot may have been affected by the preceding shots; the jobs themselves were somewhat simpler than typical jobs; and the paces observed did not include some of the extreme values occasionally seen in actual work. Some of these factors would operate to reduce error, others to increase it.

In order to develop the various phases of the results, all of the ratings were

TABLE 2

CONDENSED ANALYSIS OF VARIANCE OF PACE RATINGS

Source	Sum of Squares	Degrees of Freedom	Variance	F*
Jobs	368	3	123	48
Paces	312	4	78	31
Workers	1,456	4	366	142
Raters	4,995	5	999	381
Job×Worker	195	12	16.2	6.4
Job×Rater	559	15	37.3	15
Job×Order	36	3	12	4.7
Pace×Worker	93	16	5.8	2.3
Pace×Rater	562	20	28.1	11
Worker×Rater	155	20	7.8	3.0
Job×Rater×Order	152	15	10.1	4.0
Residual	2,270	889	2.55	
Total	12,151	1,199		

* All significant at .01 level.

converted into "errors" by subtracting from the rating the "correct" rating for that tempo. The correct rating was the ratio of the observed tempo to the normal tempo.

To facilitate handling the data, the errors were coded by dividing them by five. *The analysis-of-variance tables are in terms of this coded error.*

Differences in Level of Rating

To measure the significance of contributions to the total variance, the mean error for each rater, worker, job, pace, order of showing, and interactions was calculated. The differences among means were tested by an analysis of variance. The factors which were significantly different at better than the .01 level of confidence are shown in Table 2. Order refers to the first or second rating session.

Differences in mean ratings of raters, workers, paces. Three factors are outstanding in their contribution to total error: differences among raters, differences among workers, and differences among paces. The size of the variation among mean values of these groups is shown in Fig. 1.

Over one-third of the total variance in Table 2 is caused by rater-to-rater differences. These differences are very practically significant. For example, on a standard set by rater 6, a worker would have to perform at a pace almost 30 per cent greater than on a standard set by rater 2. These six raters were not all from the same plant, but they were all using the same verbal concept of a normal pace.

The differences in normal times which would result from time studies made on different workers are also of very practical significance. Worker 2, for example, stands out on Fig. 1 as being rated between 10 and 15 points higher than the others. No reason is advanced for this in terms of his physique. A normal time based on his performance would be 15 per cent slower than one based on worker 4.

The trend to underrate high paces and overrate low paces is also marked, and shows clearly on Fig. 1. Probably this tendency becomes even more exaggerated

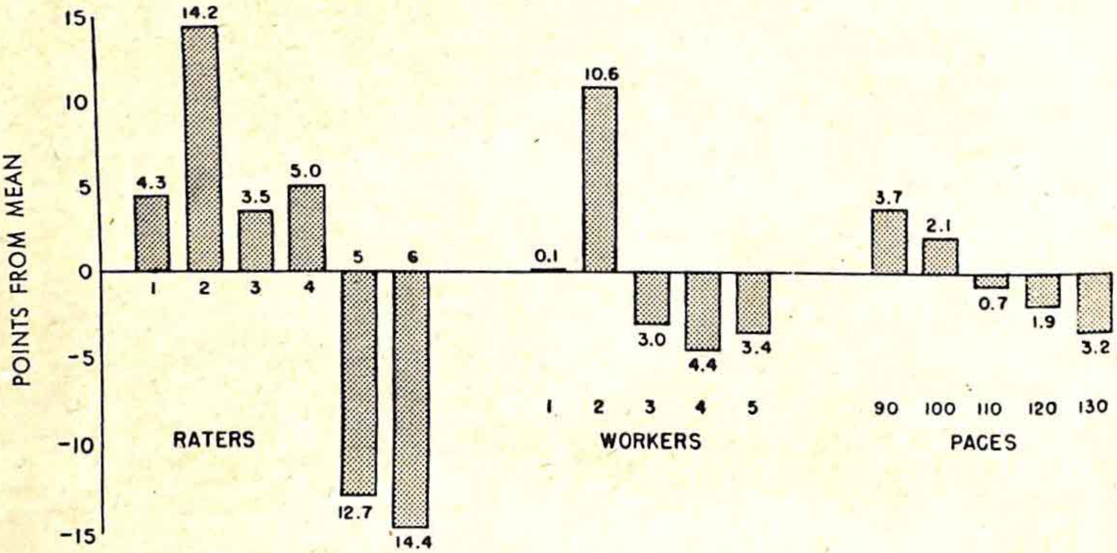


FIG. 1. Magnitude of Variation among Different Sources.

at more extreme paces, and forms some basis for workers' preferences to have time studies taken on slow operators. A standard set on an operator working at a pace of 90 would be 7 per cent "looser" than a standard set on an operator working at a pace of 130.

Differences in mean rating of jobs. Had the ratings not changed from the preliminary film to the actual study, no job differences should have occurred. Normal pace for each job was taken as the average of the raters' concepts of normal pace for that job on the preliminary film. The significant difference among job means shown in Table 2 is a result of the ratings for job 2 (stamping) being raised about 5 points from the preliminary to the actual study. In the preliminary study the raters saw the stamping from in front of the worker. They saw his hand moving at them. In the actual study the raters saw the worker from the side. This may account for the change, and also indicates that the angle of observation could be a topic for further investigation.

Interactions of practical significance among pace ratings. The significant job × worker interaction in Table 2 reveals a pattern superimposed upon the tendency for some workers to be rated higher than others and the tendency for jobs to be rated differently. When the mean worker ratings are equated and the mean job ratings are equated, there remains a tendency for some workers to be rated higher on some jobs than on others. The size of the standard deviation due to this interaction is about two rating points.

The significant job × rater and worker × rater interactions of Table 2 reveal a pattern superimposed upon the tendency for some raters to rate higher than others and the tendencies for jobs and workers to be rated differently. There remain tendencies for some raters to rate some jobs higher than others and to rate some workers higher than others. The sizes of the standard deviations due to these interactions are about 4 and 2 rating points, respectively.

Pace × rater interaction of ratings. The pace × rater interaction in Table 2 indicates that not all raters underrate the high paces and overrate the low paces.

Figure 2 shows that raters 2, 5, and 6 exhibit the tendency markedly, while rater 4 does so only slightly, and raters 1 and 3 not at all. The size of the standard deviation due to this interaction is about 5 rating points.

Other interactions of pace ratings. The pace × worker interaction in Table 2, which indicates a differential rating of the various workers at different paces, is statistically significant, but the size of the variations is small. The interactions involving order are caused by some of the raters changing their concepts of normal on certain jobs. Although statistically significant, the size of the variations is of no practical meaning.

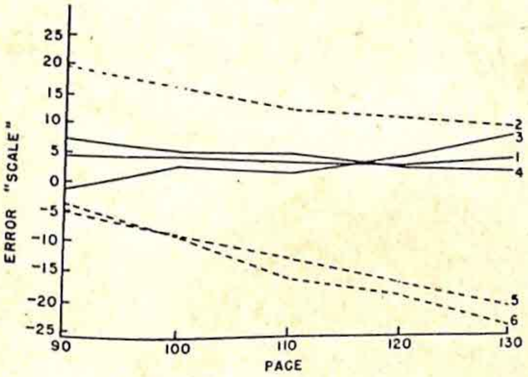


FIG. 2. Errors of Each Rater on Each Pace.

Magnitude of the Standard Errors of Pace Ratings

The standard error of all the ratings is 16 percentage points. This indicates that a normal time obtained from these ratings stands one chance in three of being wrong by ±16 points or greater.

To find the standard error representative of the consistency within a category of ratings, the variance among categories is subtracted from the total variance. The residual standard error, with all known sources of error eliminated, is 7.9 points. This represents the standard error present in any one rater's single rating.

TABLE 3

STANDARD ERRORS FOR VARIOUS POSSIBLE CIRCUMSTANCES

Item	Percentage Points
A. Gross Standard Error	16.0
B. One job, several raters rating different workers at different paces.	15.9
C. One worker on one job, rated by several raters at different paces.	15.1
D. One worker on one job at one pace, rated by several raters.	14.5
E. One rater, rating several workers at different jobs at different paces.	11.2
F. One rater on one job, rating several workers at several paces.	10.7
G. One rater rating one worker, on several jobs at different paces.	9.1
H. One rater rating one worker on one job, at different paces.	8.5
I. One rater rating a worker on a job at one pace.	7.9

Standard errors for various possible circumstances are shown in Table 3. The errors (except for the total) are measures of reliability rather than accuracy. For example, circumstance F might occur when a time-study man attempted to check a standard for a given job by timing several workers. The standard error of the normal times about the average of this rater would be 10.7 percentage points, but this rater's average might differ considerably from the average which other time-study men would get.

Circumstance B might occur in a company having the same job in several plants. The standard error of the normal times about the average would be 15.9 percentage points. Circumstances C, D, and H could occur in checking standards by various methods. Circumstances E and G might occur in a situation designed to test the reliability of raters.

Comparison of Reliability of Pace Ratings

The results of the study have been dis-

TABLE 4

Standard Errors of Pace Ratings

A. Standard Error in Points for Each Rater

Rater	Standard Error
1	13.2
2	11.8
3	16.2
4	10.6
5	10.4
6	10.1

B. Standard Error in Points of Ratings on Each Worker

Worker	Standard Error
1	14.7
2	17.0
3	14.2
4	14.3
5	15.0

C. Standard Error in Points of Ratings for Each Job

Job	Standard Error
1	12.8
2	17.7
3	15.3
4	16.6

cussed in terms of errors due to constant differences in levels of ratings and in terms of average standard errors. In addition, real differences in the consistencies of ratings within categories do exist. As Table 4 shows, some raters are more consistent than others, some jobs can be rated more reliably than others, some workers can be rated more reliably than others, and some paces can be rated more reliably than others.

Differences in reliability of pace ratings of raters, workers, and jobs. Table 4 lists the standard error of ratings for each rater, worker, and job. The differences shown are of practical significance. Raters 4, 5, and 6 are much better than rater 3, as indicated in Table 4A. Table 4B shows that a standard obtained on worker 3 or 4, for example, would be more reliable than one obtained on worker 2. The standard for job 2 would have to be checked more closely than the standard for job 1, according to Table 4C.

Differences in reliability of different paces. Figure 3 shows how reliability varies with pace. Normal pace is rated most reliably. The increase in standard error seems steeper at lower paces than higher. The standard error seems to increase proportionally to the increase in pace, about 0.2 per cent every 10 points of pace. While not of great practical significance within the range of paces studied, the trend has importance if extended to extreme paces.

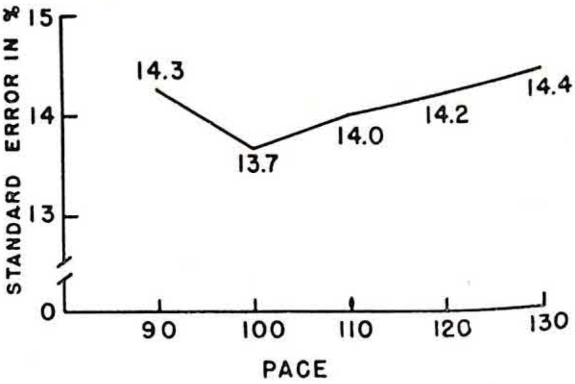


FIG. 3. Reliability of Ratings at Various Paces.

Worker Judgments

The pace judgments of each worker on each job on each pace were subjected to an analysis of variance. As with the pace ratings, these judgments were coded by subtracting the "correct" figure from them, and dividing this by five. The condensed analysis of variance, listing the factors which were significant at the .01 level of confidence, is presented in Table 5.

Worker judgments on jobs. Theoretically, normal times on different work elements or jobs should represent equally difficult levels of performance. The concept of "a fair day's work" should be con-

TABLE 5

CONDENSED ANALYSIS OF VARIANCE OF WORKER JUDGMENTS

Source	Sum of Squares	D.F.	V	F*
Paces	70	4	17.5	5.37
Jobs	142	3	47.4	22.4
Worker×Job	1,165	16	72.8	14.5
Residual	900	276	3.26	
Total	2,277	299		

* All significant at .001 level.

stant from job-to-job to both the time-study man and the worker. Other things being equal, a worker paid on an incentive plan should be able to earn as much on one job as on any other.

The results show that the workers do not equate the four jobs as the time-study men do. For example, the workers believe that "normal" on packing (job 4) is worth 12.7 cents an hour more than "normal" on stamping (job 2). Table 6 shows the differential, in points, between the workers' judgments and the pace ratings. Because the workers were given a concept of normal, this table does not indicate that the workers' concept of normal, averaged over the jobs, agrees with the raters' concept. It shows that a differential exists to the workers where none exists to the raters.

TABLE 6

DIFFERENCE BETWEEN WORKER JUDGMENTS AND PACE RATINGS

Job	Worker Judgments Minus Pace Ratings, in Points
1	−3.2
2	−6.0
3	+2.5
4	+6.7

Worker judgments of different paces. The workers exhibit the same tendency as the raters to overrate low paces and underrate high paces. Table 7 lists the mean errors on each pace.

TABLE 7

MEAN ERRORS OF WORKER JUDGMENTS ON EACH PACE

Pace	Error in Points
90	+2.5
100	+1.9
110	+1.5
120	−2.9
130	−3.0

Worker × job interaction of worker judgments. Not all workers judge the jobs in the same pattern. When the mean worker judgments and the mean job judgments are equated, there remains a tendency for some workers to judge some jobs higher than others. Table 8 shows the relative judgment made by each worker on each job. The numbers are the points from the average which the worker assigned the job, minus the sum of the points from the average of the job and the worker. The size of these interactions is of practical significance, particularly workers 1 and 4 on jobs 1 and 3. Worker 1 thinks that job 3 is worth much more than job 1, while worker 4 has the reverse opinion. On job 1, worker 1 would be satisfied with a standard time about 30 per cent less than the standard time worker 4 would require. On job 3,

TABLE 8

WORKER×JOB INTERACTION OF WORKER JUDGMENTS

(Points Deviation from the Average for the Worker and the Job)

Worker	Job			
	1	2	3	4
1	−11.8	−6.6	+17.0	+1.5
2	− 0.7	+3.4	− 2.3	−0.4
3	− 1.8	+1.4	− 2.0	+2.3
4	+17.8	+4.0	−21.4	−0.5
5	− 3.5	−2.3	+ 8.6	−2.8

worker 1 would look for a standard time about 40 per cent greater than the standard time with which worker 4 would be satisfied.

Standard error of the worker judgments. The standard error of the worker judgments is comparable to the standard error of the pace ratings. The total standard error is 13.8 points, compared to 16.0 for the pace ratings. The residual standard error is 9.0 points, compared to 7.9 for the pace ratings.

Reliability of worker judgments. Real differences in reliability of judgments exist. Some workers are more consistent than others, and some jobs are judged more reliably than others.

TABLE 9

STANDARD ERRORS OF WORKER JUDGMENTS

A. Standard Error in Points for Each Worker

Worker	Standard Error
1	15.5
2	8.2
3	7.4
4	15.3
5	15.0

B. Standard Error in Points for Each Job

Job	Standard Error
1	15.0
2	8.3
3	18.3
4	9.7

Table 9 lists the standard error of the judgments of each worker, and for each job. The differences shown are of practical significance. Some of the workers are far more reliable than others, as indicated in Table 9A. The standard error of workers 2 and 3 is half that of the rest, and compares favorably with the standard error of the time-study men. There is no apparent relation between the worker's being consistent in his judgments and the reliability with which he is rated.

The workers do not agree as well on some jobs as on others. The jobs in Table 9B on which the worker judgments are relatively unreliable are not the same jobs on which the pace ratings are relatively unreliable.

The difference in variance among paces is not statistically significant. The standard errors of the judgments seem to increase as pace increases; this was true also of the standard errors of the pace ratings.

Correlation between pace ratings and worker judgments. An analysis of the covariance between ratings of the five workers on the four jobs and the judgments of the workers about the jobs reveals a correlation within the workers and jobs of + .46. The analysis of covariance statistically equates the mean job ratings and worker judgments. The positive correlation indicates that, to the extent indicated by the correlation of .46, a worker will be rated relatively high on a job which he feels *should* be rated relatively high.

IV. PREVIOUS STUDIES OF RATING

Several experimental investigations of time-study rating have been published and are pertinent to the present study. Two have shown that filmed performances are rated the same as live performances. Other studies have attempted to measure the consistency of ratings of work pace, but in general they do not consider the effects of different raters, workers, jobs, paces, and interactions. Some investigators have shown that rating ability is improved by training, and others have been concerned with predicting rating ability. One group of researchers has been interested in the development of aids for rating, in the form of films of known paces.

Use of Film for Rating

The results of the present study would be invalid if performances on movie film are rated differently from live performances. Mundel and Margolin (8) and Barnes (1) have shown that ratings from movies are as consistent and accurate as ratings of actual workers.

Measurements of Consistency

Although few experimenters report the standard errors of the ratings they have obtained, their data can be manipulated to find the standard errors. Thirteen points represents the mean and the mode of all of the standard errors of ratings of work paces found in the studies cited in the References section. This is compatible with the findings of this study, that the total standard error is 16 points, because the previous studies have measured the reliabilities of single raters, or ratings on single jobs. None of them has had ratings on more than one worker on a job. All but one of the reported standard errors are within the range of from 8 points for residual to 16 points for total standard error reported in the present study.

Training Raters

Several investigations have been made as to whether rating can be improved by training and experience. One of Mundel's students (7) found that raters with over six months' experience had a standard error which was ⅚ the size of that of raters with less experience.

Barnes (1) cites three studies of the rating of walking, wherein the standard error was reduced from 16 points to 12 points following a two-day training session. During a six-month training period, in which a group of raters was trained to rate the specific job on which they were tested, standard error went as low as 3 points. The raters could not be expected to maintain this level of consistency on other jobs.

Lifson (4), with time-study students rating some films of factory jobs, reduced the standard error from 11 to 9 points in an eleven-week training period. He also noted that the raters who were originally most consistent improved the most. Another finding was that the concept of normal pace can be shifted very easily. He changed the level of rating by demonstrating the level of activity which was to be considered as normal, and the ratings shifted as the demonstrated level shifted.

Predicting Rating Ability

Lifson (4) found that ability to rate after training could be predicted from ability to rate before training. The correlation between a consistency measure before and after was +.45. He observed that ability to rate can be predicted even more successfully after the raters have practiced for about two weeks. Holder (3) found no correlation between rating ability and scores in the Purdue Adaptability Test, the Purdue Physical Science Test, the Purdue Mathematics Training Test, the Purdue Placement Test in English, A.C.E., or motion- and time-study tests.

Aids for Rating

In an attempt to objectify the rating procedure, Dr. M. E. Mundel has worked with a system of rating that involves the use of aids in the form of movies of established levels of pace. The rater compares the pace in question to the established standard. He attempts to compare acceleration alone, and later adjusts for job difficulties. A paper by Mundel and his students (7) reports that by using a single standard film, representing a normal pace, the standard error (with no job or worker variance) was reduced from the "unaided" value of 14 points to 13 points. Use of a series of twelve established paces, ranging from 80 to 155, further reduced the standard error to 12 points.

V. IMPLICATIONS

From the pace ratings of experienced time-study men, it was determined that the magnitude of the standard error of these pace ratings is 16 percentage points, and that much of this error is due to differences among raters, workers, jobs, paces, and their interactions.

The residual standard error is 7·9 percentage points. It has also been shown that there are real differences in the reliability of ratings from different sources, that what are two equal paces to time-study men may not feel equal to the workers performing at the paces.

Some further researches and procedures are required to help increase the consistency and accuracy of pace rating. Some specific areas of attention which are suggested by this study are discussed below.

Reliability of Raters

Some raters are more reliable than others. More research is needed in the prediction of worker 1 would look for a standard time about 40 per cent greater than the standard time with which worker 4 would be satisfied.

Standard error of the worker judgments. The standard error of the worker judgments is comparable to the standard error of the pace ratings. The total standard error is 13.8 points, compared to 16.0 for the pace ratings. The residual standard error is 9.0 points, compared to 7.9 for the pace ratings.

Reliability of worker judgments. Real differences in reliability of judgments exist. Some workers are more consistent than others, and some jobs are judged more reliably than others.

TABLE 9

STANDARD ERRORS OF WORKER JUDGMENTS

A. Standard Error in Points for Each Worker

Worker	Standard Error
1	15.5
2	8.2
3	7.4
4	15.3
5	15.0

B. Standard Error in Points for Each Job

Job	Standard Error
1	15.0
2	8.3
3	18.3
4	9.7

Table 9 lists the standard error of the judgments of each worker, and for each job. The differences shown are of practical significance. Some of the workers are far more reliable than others, as indicated in Table 9A. The standard error of workers 2 and 3 is half that of the rest, and compares favorably with the standard error of the time-study men. There is no apparent relation between the worker's being consistent in his judgments and the reliability with which he is rated.

The workers do not agree as well on some jobs as on others. The jobs in Table 9B on which the worker judgments are relatively unreliable are not the same jobs on which the pace ratings are relatively unreliable.

The difference in variance among paces is not statistically significant. The standard errors of the judgments seem to increase as pace increases; this was true also of the standard errors of the pace ratings.

Correlation between pace ratings and worker judgments. An analysis of the covariance between ratings of the five workers on the four jobs and the judgments of the workers about the jobs reveals a correlation within the workers and jobs of + .46. The analysis of covariance statistically equates the mean job ratings and worker judgments. The positive correlation indicates that, to the extent indicated by the correlation of .46, a worker will be rated relatively high on a job which he feels *should* be rated relatively high.

IV. PREVIOUS STUDIES OF RATING

Several experimental investigations of time-study rating have been published and are pertinent to the present study. Two have shown that filmed performances are rated the same as live performances. Other studies have attempted to measure the consistency of ratings of work pace, but in general they do not consider the effects of different raters, workers, jobs, paces, and interactions. Some investigators have shown that rating ability is improved by training, and others have been concerned with predicting rating ability. One group of researchers has been interested in the development of aids for rating, in the form of films of known paces.

Use of Film for Rating

The results of the present study would be invalid if performances on movie film are rated differently from live performances. Mundel and Margolin (8) and Barnes (1) have shown that ratings from movies are as consistent and accurate as ratings of actual workers.

Measurements of Consistency

Although few experimenters report the standard errors of the ratings they have obtained, their data can be manipulated to find the standard errors. Thirteen points represents the mean and the mode of all of the standard errors of ratings of work paces found in the studies cited in the References section. This is compatible with the findings of this study, that the total standard error is 16 points, because the previous studies have measured the reliabilities of single raters, or ratings on single jobs. None of them has had ratings on more than one worker on a job. All but one of the reported standard errors are within the range of from 8 points for residual to 16 points for total standard error reported in the present study.

Training Raters

Several investigations have been made as to whether rating can be improved by training and experience. One of Mundel's students (7) found that raters with over six months' experience had a standard error which was 5/6 the size of that of raters with less experience.

Barnes (1) cites three studies of the rating of walking, wherein the standard error was reduced from 16 points to 12 points following a two-day training session. During a six-month training period, in which a group of raters was trained to rate the specific job on which they were tested, standard error went as low as 3 points. The raters could not be expected to maintain this level of consistency on other jobs.

Lifson (4), with time-study students rating some films of factory jobs, reduced the standard error from 11 to 9 points in an eleven-week training period. He also noted that the raters who were originally most consistent improved the most. Another finding was that the concept of normal pace can be shifted very easily. He changed the level of rating by demonstrating the level of activity which was to be considered as normal, and the ratings shifted as the demonstrated level shifted.

Predicting Rating Ability

Lifson (4) found that ability to rate after training could be predicted from ability to rate before training. The correlation between a consistency measure before and after was +.45. He observed that ability to rate can be predicted even more successfully after the raters have practiced for about two weeks. Holder (3) found no correlation between rating ability and scores in the Purdue Adaptability Test, the Purdue Physical Science Test, the Purdue Mathematics Training Test, the Purdue Placement Test in English, A.C.E., or motion- and time-study tests.

Aids for Rating

In an attempt to objectify the rating procedure, Dr. M. E. Mundel has worked with a system of rating that involves the use of aids in the form of movies of established levels of pace. The rater compares the pace in question to the established standard. He attempts to compare acceleration alone, and later adjusts for job difficulties. A paper by Mundel and his students (7) reports that by using a single standard film, representing a normal pace, the standard error (with no job or worker variance) was reduced from the "unaided" value of 14 points to 13 points. Use of a series of twelve established paces, ranging from 80 to 155, further reduced the standard error to 12 points.

V. IMPLICATIONS

From the pace ratings of experienced time-study men, it was determined that the magnitude of the standard error of these pace ratings is 16 percentage points, and that much of this error is due to differences among raters, workers, jobs, paces, and their interactions.

The residual standard error is 7.9 percentage points. It has also been shown that there are real differences in the reliability of ratings from different sources, that what are two equal paces to time-study men may not feel equal to the workers performing at the paces.

Some further researches and procedures are required to help increase the consistency and accuracy of pace rating. Some specific areas of attention which are suggested by this study are discussed below.

Reliability of Raters

Some raters are more reliable than others. More research is needed in the prediction of reliability, and also in what it is that causes some raters to be reliable. Because of the wide differences in reliability shown among the raters in this study, the need for testing experienced time-study men is revealed. More research is also needed in methods of training time-study men to rate consistently.

Agreement of Raters

Some raters rate higher than others, even when all raters have the same verbal concept of normal pace. Agreement among raters is a major problem. It is apparent that time-study men need concentrated training in recognizing a common concept of normal pace, and frequent checks to assure that their concepts have not changed. Research is needed into methods of objectifying the rating function and into methods of securing agreement among raters.

Averaging of Several Ratings

Consideration should be given to the fact that averages are more accurate and reliable than individual ratings. Although it is expensive to have more ratings made, it is probably less expensive than setting standards which may be 30 per cent off.

Differences in Ratings on Various Workers

Some workers are rated higher than others, even when performing the same jobs in the same time, and some are rated more reliably. No research has been done, though a great deal is needed, into why some workers are rated differently from others. The reliability and level at which diffierent workers are usually rated should be known to time-study men. When there is a choice, they should have as the operator a worker who is rated reliably and at an average level.

Reliability of Different Jobs

Some jobs are rated more reliably than others. Increased reliability may come from knowing why.

Paces Rated Differently

Normal paces are rated most accurately and consistently. When there is a choice, time-study men should avoid timing extremely fast or slow operators. They should look for an operator whose performance is very close to their idea of normal.

Reliability and Accuracy of Worker Judgments

In order for any time-study system to work, paces which the time-study men rate as equal must seem equal to the workers. Time standards are not consistent from job to job unless the workers believe the standards are consistent. This study shows that this belief is not always held.

Much more can be done with worker judgments. They should be used, in some way, as the criterion with which to test the accuracy and consistency of time-study ratings. They should be used to determine adjustments and allowances. This study shows that they are reliable enough to be used for these purposes.

Some of the workers judge more consistently than the time-study men rate. This suggests the possibility of obtaining ratings from certain workers, or of having the time-study men rate while they themselves are performing the jobs.

VI. SUMMARY AND CONCLUSIONS

In a study designed to determine the nature of the errors involved in time-study ratings of work pace, six expert time-study men made ratings of the filmed performances of five workers doing each of four jobs at each of five previously established paces. The variance of the ratings was analyzed to reveal how much of the total error was attributable to differences among the time-study men in their concepts of normal work pace, differences in the way each worker and each job were rated, and differences in the rating errors made at each of the paces. Also, the standard errors of each group of ratings were compared, as a measure of relative reliability.

While the workers were performing the jobs, they evaluated, in relative terms, each pace on each job. Analysis of variance was applied to these judgments to determine the significance of the errors involved. Also, the relative level of the worker judgments on each job was compared with the relative level of the time-study men's pace ratings, to determine the amount of discrepancy.

The following conclusions are presented:

1. Pace ratings involve considerable error. The total standard error of pace ratings from different raters, on different jobs, workers, and paces was 16.0 percentage points. The standard error of one rater rating one worker on a job at a pace was 7.9 points, on the average.

2. Some raters rate higher than others. The highest rater averaged 28 points above the lowest rater.

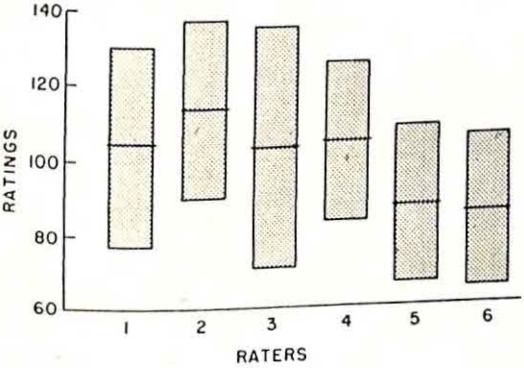


FIG. 4. Mean and "Range" of Ratings by Each Rater of the Average Normal Pace. (The "range" includes 95 per cent of the ratings by each rater.)

3. Some raters are more consistent than others. The most consistent rater had a standard error ⅝ the size of the standard error of the least consistent rater. Figure 4 shows, for each rater, the spread that would include 19 out of 20 ratings of the average normal pace.

4. Some workers are rated higher than others, even when all perform the same jobs at the same paces. A standard time set on the worker who was rated the highest would have been 15 per cent longer than one set on the worker who was rated lowest.

5. Some workers are rated more reliably than others. The standard error of the ratings on the most reliably rated worker was ⅚ the size of the standard error of the ratings on the least reliably rated worker.

6. The raters tend to overrate low paces and underrate high paces. A standard time set on a pace of 90 would have been 7 per cent longer than a standard time set on a pace of 130.

7. Normal pace is rated most reliably. The standard error increased slightly at paces slower or faster than normal.

8. Some jobs are rated more reliably than others. The standard error of the ratings on the most reliably rated job was ¾ the size of the standard error of the ratings on the least reliably rated job.

9. Interactions are important. Some raters rated some jobs higher than others; some raters rated some workers higher than others; some workers were rated higher on some jobs than on others; and not all raters followed the pattern of underrating high paces and overrating low ones.

10. Workers' judgments on equating the jobs differ from the pace ratings of the time-study men. The workers believed there was as much as a 12 per cent per hour differential between two "normal" paces as established by time-study men.

11. Individual differences among the worker judgments are very important. On the same job there was as much as a 40 per cent difference between two workers' judgments as to what was a normal pace.

12. Some workers can judge more reliably than time-study men can rate. The standard error of the judgments of the two most consistent workers was ¾ the size of the standard error of the best time-study men.

13. A correlation of +.46 exists between the workers' judgments and the pace ratings of the time-study men. To

the extent indicated by this correlation, on a job which a worker judged relatively high, he would be rated relatively high; and on a job which a worker judged relatively low, he would be rated relatively low.

In conclusion, this study has indicated the nature of the errors which are involved in time-study pace ratings, and has indicated some possible approaches to the reduction of these errors.

REFERENCES

1. Barnes, R. M. *Work measurement manual.* (3rd Ed.) Dubuque: Wm. C. Brown Company, 1947.
2. Edwards, A. L. *Experimental design in psychological research.* New York: Rinehart, 1950.
3. Holder, W. B. Applicability of six standard tests in the prediction of time study rating ability. Unpublished master's thesis, Purdue Univer., 1949.
4. Lifson, K. A. Performance rating. *Time Study Engr,* 1951, **6** (June), 179-181.
5. Lifson, K. A. A psychological approach to pace rating. Unpublished doctor's dissertation, Purdue Univer., 1951.
6. Lynch, H. R. Rating of time studies. In *Proceedings of the 5th annual time study and methods conference.* 1950.
7. Mundel, M. E. (Ed.) *Report of the 5th annual motion and time study work session.* Lafayette: Purdue Univer., 1950. (Mimeographed)
8. Mundel, M. E. & Margolin, L. *Report of the 4th annual motion and time study work session.* Lafayette: Purdue Univ., 1948. (Mimeographed)
9. Schell, H. A. A study on effort rating. *Modern Mgmt,* 1949, **9** (April), 19-20.

(Accepted for publication January 23, 1953.)

Vol. 67, No. 6 Whole No. 356, 1953

Psychological Monographs: General and Applied

A Quantification, Standardization, and Validation of the Bender-Gestalt Test on Normal and Neurotic Adults[1]

Wallace Gobetz
New York University

I. PURPOSE

In recent years there has been an increasing emphasis on the factor of motivation as an influence on the perceptual processes. A perception is not a literal copy of the stimulus, and individual differences in perceiving cannot be explained solely in terms of the objective stimulus pattern, primary configural tendencies, or idiosyncrasies in sensory, neural, or motor capacities. Needs, wants, values, and tensions can be significant organizing factors in perception; and it follows that the more unstructured, ambiguous, and amorphous the stimulus situation, the greater the likelihood that any given individual will project his needs into the situation in an effort to give it meaning and form, and the greater the individual variation in perceptual response that may be anticipated.

Although the interest of the clinical psychologist and the student of personality developed some time after the experimental investigation of perception in the laboratory, the diagnostic and therapeutic possibilities of graphic-motor behavior as revealed in artistic productions were recognized by the clinician long before such behavior was made the object of systematic experimental investigation. Anastasi and Foley (1, 2, 3, 4) in a series of articles surveyed the literature on artistic behavior in the abnormal, a literature in which theory and speculation are almost as numerous as undisputed fact. Interest in graphic behavior has not, of course, been limited to the study of art and art forms, since it has been recognized in the past two decades that virtually all forms of graphic expression may have implications for personality evaluation and diagnosis. Relevant in this connection are the approaches of Werner Wolff (46, 47, 48, 49, 50), Saudek (38), Machover (30), Mira (31), Sapas (36), Foster (22), and Bender (9, 10, 11, 12, 13, 14, 15).

Bender is generally given credit for taking the experiments on the reproduction of visually perceived form, whether from copy or from memory, out of the laboratory and into the clinical setting, and there is little question that she did more than any other individual to see the diagnostic possibilities of the method and popularize its use. Impressed by the experimental studies of Wertheimer, Köhler, and Koffka in the Gestalt theory of perception, she published a monograph (12) in 1938 summarizing her findings on the alteration of the "Gestalt function" in the copying of geometrical designs by children and abnormal patients. The test itself consists of nine geometrical designs which were selected from approximately thirty which Wertheimer (44) used in his classical study of Gestalt principles. Each design is on a separate card, and the subject is asked to copy them one at a time. Other reports of the clinical use of the Bender-Gestalt Test have been published by Schilder (39, 40), Bender, Curran, and Schilder (16), Orenstein and Schilder (34), Fabian (20), Hutt (27, 28), Wayne, Adams, and Rowe (41), Harrower (25), David Wechsler (42), Israel Wechsler (43), Barkley (7), Billingslea (17), Woltmann (51), and Pascal and Suttell (35).

---

[1] This study is an abridgment of a dissertation in the Department of Psychology submitted to the faculty of the Graduate School of Arts and Science, New York University, June, 1952, in partial fulfillment of the requirements for the degree of Doctor of Philosophy. The writer wishes to express his gratitude to Professor Thomas N. Jenkins, whose perspicacious observations gave impetus to the exploration of promising leads, and to Dr. William D. Glenn, Jr., former director of the New York University Testing and Advisement Center, who very kindly consented to the special testing of the World War II veteran and nonveteran clients. Heartfelt thanks are likewise proffered to Dr. Bernard Locke, chief clinical psychologist of the Mental Hygiene Unit of the Brooklyn Regional Office of the Veterans Administration, and to Dr. Richard H. Paynter, chief clinical psychologist of the Mental Hygiene Unit of the New York Regional Office of the Veterans Administration, whose generous cooperation made the testing of neurotic outpatients possible. Finally, the writer wishes to acknowledge his great debt to his wife, whose devotion and warm understanding served as a never-ending source of inspiration and to whom this study is affectionately dedicated.

While Bender was primarily interested in the Gestalt test as an instrument for the exploration of the Gestalt function, with emphasis upon its maturational and pathological aspects, Hutt (28) conceived of the test as a projective personality technique, rich in interpretive significance and psychodynamic implications and capable of contributing valuable clues to differential diagnosis. Clinical psychologists have been quick to follow Hutt's lead rather than Bender's more prosaic approach because they have long sought for a projective device less time-consuming than the Rorschach or TAT which may be employed as a brief supplement to the longer techniques in yielding new or confirmatory personality data. Moreover, Bender's monograph is of rather limited usefulness to the clinician because of its emphasis on extreme deviations, the relative absence of control data, and the failure to support interpretations with quantitative findings. The reader is generally left in the dark concerning the actual number of cases upon which she bases her conclusions. One gets the impression that there is a Bender-Gestalt record "typical" of the schizophrenic or "typical" of the aphasic and is then puzzled when he runs across a record in the clinic which fails to fit the "classical" picture. But if Bender's articles lack specificity, the criticism might be leveled against Hutt that he offers "too much too soon" and that claims are made for the test which have yet to be validated. Although he states that the "clinical syndromes" for the psychoneurotic, the schizophrenic, and the brain-injured patients are "substantiated by the author's research studies," he has not (to the writer's knowledge) published his original data in quantitative form, by way of reporting either validity coefficients or the proportion of false positives or false negatives to be anticipated. Furthermore, many of his "determinants," or scoring deviations, are vaguely defined, resulting in varying interpretations of them by clinicians who employ the same terms but assign them different meanings.

It may well be that Bender and Hutt are capable of drawing inferences from the test which provide valuable insights into the patient with whom they are working, but it may also be that they are relying heavily upon minimal cues based upon long experience with the test which are generally labeled as "clinical intuition" and which may be less readily communicable to others. The fact of the matter is that the majority of clinical psychologists using the test interpret it in a more or less rough, subjective, global manner, drawing psychodynamic implications on the basis of essentially unsystematized clinical observation. If the test is to have real value to the clinician and if it is to be given univocal interpretation by those who use it as a diagnostic or prognostic instrument, both Bender's and Hutt's provocative ideas should be stated as hypotheses to be subjected to rigorous experimental or clinical validation.

The need for quantification and validation of all projective tests has been recognized with increasing frequency in recent years, despite the objections of some clinicians who seem to feel that any attempt to apply statistics to projective devices would automatically destroy their usefulness as instruments for assessing the "global personality." Sargent (37) has presented a strong case for quantification, and White (45), Bell (8), Cronbach (18), and Murphy (32) have expressed themselves in a similar vein. With regard to the Bender-Gestalt Test, the majority of psychologists using the test seem to feel that it is helpful in working with organics and psychotics but disagree as to its validity with neurotics. Bender (12) reports no data on neurotics, and Woltmann notes that "very often . . . a clearly established diagnosis of psychoneurosis from the test battery is accompanied by very normal copies of the Gestalt figures" (51, p. 348). Billingslea (17) and Pascal and Suttell (35) have made efforts to quantify the Bender-Gestalt, arriving at almost diametrically antithetical findings with regard to the neuroses. Billingslea's findings are essentially negative, while Pascal and Suttell's are strongly positive.

The general purpose of the present investigation is to determine whether neurotics and normals can, in fact, be distinguished on the basis of their Bender-Gestalt records. Most clinical psychologists are agreed that psychotics and organics show gross deviations on the test that are sufficiently bizarre or unusual to identify such cases fairly readily without the necessity of developing an elaborate scoring system, but, as noted above, are less certain of the test's applicability to patients suffering from neurotic conditions. Before inquiring into the validity of the psychodynamic interpretations of the individual scoring factors as applied to neurotics, it would appear that the logically prior step would be to ascertain whether the scoring factors themselves differentiate between neurotic and normal subjects. If a sign alleged to reveal anxiety or emotional immaturity occurs with equal frequency in both groups, one might well question the validity of such an interpretation. The specific aims of this investigation, then, are to ascertain the discriminating power of the various scoring elements, to isolate any which exhibit differential validity as regards normals and neurotics, and, if possible, to combine these discriminating elements into a total score which will effectively separate the contrasting clinical groups. To accomplish these aims, it is proposed to develop an objective scoring system, defining the scoring elements operationally rather than conceptually or interpretively, and to subject the observed deviations of reproductions from stimulus figures to a statistical appraisal similar to that used in item-analysis techniques. It should be emphasized that this is a study of empirical validity; the effort is not to identify psychological processes or psychodynamic correlates, but to define the test's uses and limitations in practical terms.

II. PROCEDURE

A. Administration of the Test

Although Bender did not provide any standard instructions for administering the test, almost all users of the Bender-Gestalt have employed the same basic method of administration. The stimulus cards, each of which contains a test figure, are presented one at a time, and the subject is asked to copy them while keeping the card in full view. Rulers or mechanical guides are not permitted, but otherwise the subject is free to proceed as he chooses. Questions about how the figures are to be copied are usually answered in a noncommittal manner or referred back to the subject for decision. There is no time limit, and the reproductions themselves are not timed. Erasures, crossing out of part of a figure, or making more than one attempt to reproduce a figure are allowed.

In the present study three major innovations in administration of the test were introduced. (*a*) Each reproduction was timed. Although all previous investigators have ignored the time factor except to observe that extremely long or short times should be noted as a qualitative datum, the present investigator felt that time might be a significant factor in its own right and that its inclusion was as justifiable in connection with the Bender-Gestalt as it has been with word association tests and the Rorschach. (*b*) The administration of the initial test was followed by a test of immediate recall. As soon as the subject had copied all the figures, all stimulus cards were removed from sight and he was asked to draw the figures from memory. This modification of the basic procedure was introduced on the hypothesis that if any differences in test performance are exhibited by normals and neurotics, they will be accentuated in the absence of an objective stimulus. Numerous studies (e.g., 21, 23, 24) on memory for form show pronounced alterations of the reproductions in normal adults, and it is conceivable that certain latent tendencies that are held in check during the copying process might, under the pressure of an unexpected test of incidental memory, reveal themselves as distinctive of neurotic adults. Moreover, if emotional blocking is heightened under these conditions, one would expect the total number of figures recalled by the neurotic group to be significantly smaller than for the control group. (*c*) An immediate retest followed the immediate recall. As soon as the subject had completed his recall of the figures, he was asked to copy the figures from the models in a procedure identical with that employed for the initial test. This procedure permits a comparison of performance under conditions in which the test materials and procedures are first unfamiliar and then familiar and sheds light on the question of whether observed deviations occurring during the first administration of the test persist or disappear under a second administration. Unlike previous studies, such as that of Pascal and Suttell (35), in which retests were administered to a comparatively small proportion of the total group tested merely for purposes of estimating reliability, the retest was employed in the present research as an integral part of the testing procedure with implications for both validity and reliability.

The test was administered to the control group by psychologists at the New York University Testing and Advisement Center and to the experimental group by clinical psychologists at two Veterans Administration mental hygiene units in New York City. While the writer administered a number of the tests himself, the great majority of tests were given by others on the assumption that the results would have more universal applicability if obtained under conditions approximating those in actual clinical practice. This procedure would also reduce to a minimum any unconscious bias on the part of the present investigator to influence the outcome. The instructions were explained and illustrated by the writer at staff meetings of the installations where the test was given, and opportunities were provided for the giving of practice tests to answer any questions that might arise.

B. SUBJECTS

The neurotic group consisted of 108 World War II veterans who, at the time of testing, were undergoing treatment at the mental hygiene service of the Veterans Administration. The cases were drawn from the mental hygiene clinics of the New York Regional Office and the Brooklyn Regional Office, with the majority coming from the latter installation. Inservice diagnoses, on the basis of which these subjects were receiving disability pensions, were not used in view of the suspicion that many such diagnoses were more a matter of practical expedience than of scientific accuracy. Rather, the nosological classifications were predicated upon judgments arrived at by one or more staff psychiatrists on the basis of case histories, clinical interviews, and the results of an extensive battery of diagnostic psychological tests administered by staff clinical psychologists. The Bender-Gestalt was incorporated into the battery for purposes of this study so that it would be regarded by the subjects as part of the normal testing procedure, but in no instance was the test used to establish the diagnosis. If there was any doubt in the minds of the psychiatrists or psychologists about the diagnosis of a given patient, the case was not included. Cases with complicating features such as gunshot wounds, ulcers, migraine, asthma, psychopathic personality, schizoid trends, orthopedic disabilities, or neurological involvement were automatically excluded, the effort being to obtain "pure" neurotic cases in so far as possible. There was less concern for differential diagnosis within the neurotic rubric than for differential diagnosis between the neurotic and nonneurotic psychiatric patients in view of the unreliability of specific diagnostic categories, which Ash (6) and others have demonstrated. Only native white male adult outpatients were included in the experimental criterion group. The criteria for selecting this group were so rigorously adhered to that over three years were required to collect the data.

The control or "normal" group consisted of 285 World War II veterans availing themselves of aptitude testing and vocational counseling under the provisions of Public Law 346, generally referred to as the GI Bill of Rights. Only native white male nondisabled veterans receiving no pension for physical or neuropsychiatric disability were included. For purposes of this study, a veteran was considered "normal" provided he had no history of neuropathic traits or emotional maladjustment, gave no evidence of nervous mannerisms or emotional disturbance during two hour-long interviews or during a series of testing sessions ranging in total time from three to eleven hours, had not been given a neuropsychiatric diagnosis in service, and had not subsequently requested or received psychiatric treatment.

In order to refine the control group still further, the subjects in this group were delimited by the application of a psychometric as well as a clinical criterion. The psychometric instrument used for this purpose was the Minnesota Multiphasic Personality Inventory (group form), which has been clinically standardized and validated. It is believed that deliberate misrepresentation of responses, a pitfall in employing any personality inventory, was reduced to a minimum by administering the test in a vocational-guidance rather than a personnel-selection situation. Since a standard score of 70 is generally regarded as the dividing line between normal and abnormal scores on any one of the nine scales, the control group was subdivided into three subgroups as follows:

1. *Below-70 Group* ($N = 155$). All MMPI scores fall below 70.

2. *1-above-70 Group* ($N = 68$). Only one of the MMPI scores equals or exceeds 70.

3. *2-above-70 Group* ($N = 62$). Two or more of the MMPI scores equal or exceed 70.

This breakdown into subgroups permits comparisons within the control group, as well as between the control group and the neurotic subjects, and increases the probability that at least one sizable portion of the controls is sufficiently "normal" to meet the rigorous standards of a research study using the method of contrasting groups to establish test validity.

Along with the MMPI, all control subjects took the Otis Gamma test so that the relationship between Bender-Gestalt Test performance and intelligence might subsequently be explored. Ideally, it would have been desirable to have administered the MMPI and the Otis to the neurotic subjects as well, but it was not feasible to make such additions to the regular diagnostic test battery employed at the mental hygiene clinics from which the experimental subjects were drawn because the need to process cases would have made the testing time prohibitive. In the neurotic group, 84 of the 108 subjects were, however, given the Wechsler-Bellevue, and the relationship between Bender-Gestalt Test performance and scores on this test will be discussed subsequently.

The mean age of the total control group ($N = 285$) is 24.22 years, while that of the total neurotic group ($N = 108$) is 29.76 years. The standard deviations are 4.70 and 5.74, respectively. With regard to education, the mean is 12.47 grades for the controls and 10.74 for the neurotics, with standard deviations of 1.72 and 2.89, respectively. Of the normals, 68.7 per cent are single, compared with 37.0 per cent of the neurotic subjects. Of the controls, 69.7 per cent are Army veterans, 23.9 per cent Navy, 3.9 per cent Marines, and 2.5 per cent Coast Guard. For the neurotics, the corresponding figures are 84.3, 14.8, 0.9, and 0.0 per cent. The three subgroups of the total control group are remarkably similar with regard to age, education, marital status, and branch of service, and it is noteworthy that the majority of subjects from both the control and neurotic groups are between 20 and 30 years of age, the modal grade is the 12th, and the modal service affiliation is the Army.

III. THE SCORING SYSTEM

Bender did not develop any system of scoring for her test because her concern with gross distortions and bizarre deviations of the reproductions from the original stimuli made such a refinement unnecessary. In the present research the working hypothesis has been adopted that subtle differences between normals and neurotics may in fact exist, and that the way to test this hypothesis is to develop a reasonably objective and communicable method of scoring that will reflect those nuances of drawing which contribute to relatively minor modifications of the original stimuli.

A distinction has been made in the present investigation between "graphic" signs and "methods" signs.[2] The graphic signs are defined as those characteristics of the reproductions which are scoreable solely by inspection or measurement of the figures appearing on the test record; they do not presuppose observation of the subject's test behavior. Examples are size, contiguity, substitution, and asymmetry. The methods signs are defined as those characteristics of the reproductions which are scoreable only by direct observation of the subject's test behavior. Examples are counting, paper rotation, direction, and initial part.

The final list of signs consists of 82 scoring categories, many of which are represented in more than one figure, the total number of global and individual figure signs being 312. Each scoring category is numbered, and an individual sign is designated by this number and by the number of the figure on which it appears. "Guides," for example, is scoring category 9; should guide lines be used to reproduce figure 5, the designation would be 9-5. Similarly, "initial part" is category 65; should the diamond in figure A be made before the circle is drawn, and should the bell of figure 4 be drawn before the open square, the designations would be 65-A and 65-4, respectively. Certain categories are represented in only one figure as, for example, "pairing" (number 70) which occurs solely in figure 1 and is designated as 70-1. Other categories, such as "rotation: total" (number 48) may occur in all figures, and the designation is 48-A, 48-1, etc. "Global" categories such as sequence, cohesion, and colliding refer to the placement of all the figures on the page and hence have no specific figure designation.

The scoring categories are defined below. It will be noted that certain numbers are omitted in designating categories (e.g., there are no categories numbered 3, 4, or 13). Originally, the individual figure categories were numbered from 1 to 80, but certain categories proved to be impracticable or too susceptible to subjective judgment in an exploratory trial run completed prior to the undertaking of the present study. Rather than renumber the retained categories, it was decided to adhere to the earlier numerical designations.

A. Individual Figure Signs

1. *Length* (all figures). Distance (in millimeters) of the horizontal extent of the reproduction. Measure from the extreme left to the extreme right of the reproduction, using a transparent millimeter ruler.

2. *Height* (all figures). Distance (in millimeters) of the vertical extent of the reproduction. Measure from the extreme upper to the extreme lower part of the reproduction, using a transparent millimeter ruler.

5. *Tremulous line* (figures A, 4, 6, 7, 8). A shaky, quivering, wavering line.

6. *Erasure* (any figure). All or part of a reproduction exhibits evidence of erasure.

7. *Multiple attempt* (any figure). Two or more reproductions of a given figure appear on the test record. The initial reproduction may involve the entire figure or just part of it. The initial reproduction may be crossed out, partially erased, or left intact. Not scored if the initial reproduction is erased completely and the second attempt is superimposed upon it.

8. *Figure A placement* (figure A). One or more reproductions appear above figure A on the test record, or figure A is three inches or more from the top of the page.

9. *Guides* (any figure). Use of extraneous lines or points as an aid to reproducing the figure.

---

[2] For definitions of the specific "graphic" and "methods" signs mentioned in this paragraph, see section A below, entitled "Individual Figure Signs."

10. *Overshooting lines* (figures A, 4, 7, 8). Lines overshoot one another at points of juncture.

11. *Retracing* (figures A, 4, 6, 7, 8). A part of a line or one or more lines are made with repetitive strokes which are more or less superimposed on one another. The line is belabored rather than sketched.

12. *Projection retracing* (figures 7, 8). One or more of the pointed projections of figures 7 or 8 are retraced as defined in category 11. Scored also for Retracing (category 11). Not scored when entire reproduction is retraced. Not scored if only the diamond in figure 8 is retraced.

14. *Substitution: dot-dash* (figures 1, 3, 5). Dots are reproduced as horizontal or vertical dashes at least $1/16$ inch in length. At least four dots should be so reproduced to be scored.

15. *Substitution: dot-circle* (figures 1, 3, 5). Dots are reproduced as unfilled circles or loops. At least four dots should be so reproduced to be scored.

16. *Substitution: dot-ball* (figures 1, 3, 5). Dots are reproduced as filled-in circles or loops. At least four dots should be so reproduced to be scored. Slightly enlarged dots are not scored.

17. *Substitution: dot-scribble* (figures 1, 3, 5). Dots are reproduced as "wiggly" lines, "butterfly" lines, or scribbles. At least four dots should be so reproduced to be scored.

18. *Substitution: circle-dot* (figure 2). The circles in figure 2 are reproduced as dots or filled-in loops. At least four circles should be so reproduced to be scored.

20. *Numeration* (figures 1, 2, 3, 5, 6). The number of dots, circles, columns, or waves in the reproduction differs from the actual number in the stimulus figure. The correct number of elements is as follows for the figures concerned:

figure 1	12 dots
figure 2	11 columns or 33 circles
figure 3	1, 3, 5, and 7 dots
figure 5	19 dots in semicircle, 7 dots in tangential line
figure 6	4 wave crests in horizontal and 4 in vertical

21. *Wave accentuation* (figure 6). Marked increase in amplitude of horizontal or vertical wavy lines. At least two of the waves should be so reproduced to be scored.

22. *Wave flattening* (figure 6). Marked decrease in amplitude of horizontal or vertical wavy lines. At least two of the waves should be so reproduced to be scored.

23. *Wave irregularity* (figure 6). Marked variation in wave length or amplitude of horizontal or vertical wavy lines. At least one of the waves should differ markedly from the others to be scored.

25. *Curvature contraction* (figures 4, 5). Marked constriction of bell in figure 4 and/or semicircle in figure 5.

26. *Curvature flattening* (figures 4, 5). Marked flattening of bell in figure 4 and/or semicircle in figure 5.

27. *Asymmetry* (figures A, 3, 4, 5, 7, 8). A symmetrical part of a figure is reproduced asymetrically. Lines equal in length are made unequal, or corresponding angles equal in size are made unequal, or regular contours are made irregular.

28. *Disproportion* (figures A, 4, 5, 6, 7, 8). Distortion in the relative size of the parts of a figure. Parts of equal length, height, or area are made unequal.

29. *Displacement* (figures A, 3, 4, 5, 6, 7, 8). Shifting of the point of juncture of one part of a figure with the other, either to the left or right or above or below. In the case of figure 5, junction of the tangential line at the center of the semicircle is scored, as well as junction at the extreme right. Scoring for figure 6 is strict; the vertical line must cross the horizontal at the middle of the third horizontal wave; displacement is also scored if the horizontal line crosses the vertical in the trough at either end of the second (from the top) vertical wave. In the case of figure 7, displacement is scored when one hexagon is drawn well below the other or when the lower point of the left hexagon is not below the lower point of the right hexagon.

30. *Contiguity* (figure A, 4, 5, 7, 8). Contiguous parts of a figure overlap or are only partially contiguous, or overlapping parts are made contiguous.

31. *Angulation* (figures 2, 3, 4, 5, 6, 7). Distortion of the angular direction of a figure or of the angle formed by one part of a figure with the other.

32. *Concentric arc* (figure 3). Two or more of the arrowheads in figure 3 are reproduced as arcs.

33. *Parallel lines* (figures 4, 7, 8). Parallel lines are reproduced as converging or diverging.

34. *Parallel columns* (figure 2). Slanting columns of figure 2 are reproduced as converging or diverging. Also scored for angulation (31-2).

35. *Parallel rows* (figure 2). The parallel rows of figure 2 are reproduced as converging or diverging.

36. *Horizontal irregularity* (figures 1, 2). The horizontal progression of dots or columns is irregular and uneven rather than correctly aligned or consistent in direction.

37. *Irregular spacing* (figures 1, 2, 3). The distances between dots, circles, rows, or columns are markedly unequal.

38. *Perseveration* (figures 1, 2, 3). Tendency to extend a figure indefinitely. There must be at least four additional dots in figure 1, four additional columns in figure 2, and one additional column in figure 3. Also scored for numeration (category 20).

39. *Serial incompleteness* (figures 1, 2, 3). Tend-

ency to reduce the number of elements in figures 1, 2, and 3. Scored when there are nine or less dots in figure 1, eight or less columns in figure 2, three or less columns in figure 3. Also scored for numeration (category 20).

40. *Closed-open* (figures A, 4, 7, 8). Enclosed figures or parts of figures are not completely closed in the reproduction because of the presence of gaps. One instance is sufficient to score.

41. *Open-closed* (figures 4, 5). Open figures or parts of figures are closed or nearly closed in the reproduction.

44. *Splitting* (figures A, 4, 7). The parts of a figure, joined in the stimulus figure, are reproduced so that they are separated by a narrow space (less than 1/4 inch).

45. *Dissociation* (figures A, 4, 5, 7). The parts of a figure, joined in the stimulus figure, are reproduced so that they are separated by a space of 1/4 inch or more.

46. *Simplification* (any figure). The reproduction is made considerably less complex than the original, usually by substitution of solid for dotted lines, straight lines for angles, loops or curved lines for angular figures, or straight lines for wavy lines. Also scored for distortion (category 47).

47. *Distortion* (any figure). Marked deviation of the reproduction from the original, including simplification, addition or omission of angles, massing of dots, marked line irregularity, closure of open figures, etc.

48. *Total rotation* (any figure). The entire reproduction is rotated 45 degrees or more from the orientation of the stimulus figure. Turning of the stimulus card or of the paper with corresponding orientation of the reproduction is not scored.

49. *Part rotation* (figures A, 4, 5, 7, 8). Rotation of 45 degrees or more of one of the parts of a figure. An exception is figure 5, which is scored for a rotation of the semicircle of 30 degrees or more.

50. *Reversal* (figures A, 4, 7). The spatial position of the two parts of a figure are reversed so that the left part is reproduced on the right and the right part on the left.

51. *Paper rotation* (any figure). The paper is turned by the subject to a horizontal position while one or more figures are reproduced. A reproduction so made is not scored for total rotation (category 48) if it is correctly oriented with respect to the horizontal orientation of the paper.

52. *Card rotation* (any figure). The stimulus card is turned by the subject at least 90 degrees in any direction. A reproduction made under these conditions is not scored for total rotation (category 48) if it is correctly oriented with respect to the orientation of the card.

53. *Fragmentation* (any figure). Only an isolated part, or fragment, of the total figure is reproduced. In the cases of figures 1 and 2, four dots or less and three or less columns are so scored; in these two instances, serial incompleteness (category 39) would also be scored.

54. *Omission* (figures A, 2, 3, 4, 5, 6, 7, 8). Omission of a major part of a figure in the reproduction. More of the figure is present, however, than would be true if fragmentation were scored.

55. *Upward slope* (figures A, 1, 2, 3, 6, 8). The reproduction inclines upward. Scoring is strict; any perceptible upward slant is scored, even if slight. The horizontal line of figure 6 is normally slanted in the stimulus card but is scored if also slanted in the reproduction, since most subjects fail to observe slant in that figure. Scored also for counterclockwise rotations up to 45 degrees.

57. *Direction: sinistrad* (figures 1, 2, 3, 4, 5, 6). Horizontally oriented stimulus figures are reproduced from right to left. In figure 4, either the open square or the bell, or both, may be made from right to left. In figure 5 only the semicircle, and in figure 6, only the horizontal line are scored.

58. *Direction: upward* (figures 2, 3, 6). A figure, part of a figure, or a line is drawn in the upward direction, i.e., from below to above. In figures 2 and 3, any one column so drawn is sufficient for scoring.

59. *Direction: centrifugal* (figures 4, 5). The bell of figure 4 or the semicircle of figure 5 is made by starting at the crest of the curve and drawing lines from the center of the curve to each end.

60. *Counting: test figure* (figures 1, 2, 3, 5, 6). Subject overtly counts the number of dots, circles, columns, rows, or waves in the stimulus figure, as indicated by moving pencil or finger over each element of the figure, counting aloud, moving lips while counting subvocally, etc.

61. *Counting: reproduction* (figures 1, 2, 3, 5, 6). The same as category 60, except that the elements of the reproduction, rather than the stimulus figure, are counted.

62. *Recounting: test figure* (figures 1, 2, 3, 5, 6). Subject overtly counts the elements of the stimulus figure, as defined in category 60, two or more times.

63. *Recounting: reproduction* (figures 1, 2, 3, 5, 6). The same as category 62, except that the elements of the reproduction, rather than the stimulus figure, are overtly counted two or more times.

64. *Rows* (figure 2). Figure 2 is reproduced row by row rather than column by column. Making the top row only and then completing the columns is not scored.

65. *Initial part* (figures A, 4, 5, 6, 7, 8). This refers to the part of the figure which is reproduced before the other part is made. Scored only if the part indicated below is made first:

figure A—diamond

figure 4—bell
figure 5—handle
figure 6—vertical line
figure 7—left hexagon
figure 8—enclosed diamond

66. *Angle differentiation* (figure 7). The two angles at the extremes of either hexagon are made essentially alike rather than as acute and obtuse.

67. *Leg shortening* (figure 5). One "leg" of the semicircle is shorter than the other. The angle formed between the horizontal and an imaginary line connecting the two ends of the legs should be at least 10 degrees. An angle of 45 degrees or more is scored for both leg shortening and part rotation (category 49).

68. *Side reduction* (figures 7, 8). A six-sided figure is made with five or less sides. Scored also for distortion (category 47).

72. *Correction* (all figures). An inaccuracy in the reproduction is rectified by drawing another line without erasing the incorrect one, by crossing out a part of the figure, or by partially retracing the incorrect part.

73. *Multiple stroking* (figures A, 4, 6, 7, 8). The subject takes two or more strokes to make a line which is usually made by one continuous stroke. Sketching is excluded from this definition, however.

74. *Projection angulation* (figures 7, 8). The sides of the angles forming the pointed projections in figures 7 or 8 are made unequal, lopsided, or asymmetrical; the points of the angles are blunted or otherwise distorted; or, in the case of figure 8 only, the corresponding angles at the extremes of the hexagon are unequal. Scoring is strict; slight deviations are scored.

75. *Line sag* (figures A, 4, 7, 8). Straight lines are reproduced as curved, sagging, convex, or concave.

77. *Curvature: horseshoe* (figure 5). The legs of the semicircle in figure 5 curve in toward each other, resembling a horseshoe.

78. *Continuation* (figures 1, 2). Figure 1 or 2 is made on two lines, usually when the reproduction is incomplete at the right-hand margin.

80. *Curve squaring* (figures 4, 5). The bell of figure 4 or the semicircle of figure 5 resembles an open square.

B. GLOBAL SIGNS

Sequence. Order in which the reproductions are placed on the page. Scored in terms of the number of sequential connections. A sequential connection refers to any two reproductions which follow one another in the same order as the presentation of the stimulus cards. A reproduction may be directly above, below, to the right, or to the left of the one which it follows. A connection exists only when no other figure intervenes between the two successive reproductions. If a figure is near the right-hand margin and is followed by its successor near the left-hand margin below, a connection is scored. If a figure is at the bottom of the left half of the page and is followed by its successor at the top of the right half of the page, or vice versa, a connection is scored. If a figure is at the bottom of a page and is followed by its successor on the obverse side of the paper or on a new sheet, a connection is scored.

Rigid sequence. Each reproduction is placed directly below the preceding one. The reproductions may all be placed on one side of the page or continued from the front side to the obverse side or to a new sheet of paper. Not scored, however, if the reproductions are placed on one side of the page in two columns. Also scored for 8 sequential connections.

Cohesion. The degree of compression or expansion of all the reproductions, expressed in terms of the amount of space filled, as follows:

Cohesion: ⅓. All of the figures are reproduced in the upper, middle, or lower third of one page.
Cohesion: ½. All of the figures are reproduced in the upper or lower half of one page.
Cohesion: ⅔. All of the figures are reproduced in the upper or lower ⅔ of the page.
Cohesion: 1. All of the reproductions occupy a full page.
Cohesion: 1⅓. Reproductions occupy all or part of one page plus ⅓ of the back of the page or ⅓ of a new sheet.
Cohesion: 1½. Reproductions occupy all or part of one page plus ½ of the back of the page or ½ of a new sheet.
Cohesion: 1⅔. Reproductions occupy all or part of one page plus ⅔ of the back of the page or ⅔ of a new sheet.
Cohesion: 2. Reproductions fully occupy the front and back of a page or one side of one sheet and one side of a second sheet.
Cohesion: 2+. More than 2 pages (front and back of one sheet or one side of each of two sheets) are used to reproduce the figures.

Second sheet. Use of more than one sheet of paper on which to place the reproductions. Refers to the use of two or more separate sheets, not to the front and back of a single page.

Collision. Two or more reproductions overlap or run into one another.

Numbering. One or more of the reproductions are numbered by the subject.

Compartments. One or more lines are drawn to separate the reproductions.

Total time. Sum (in seconds) of the times required to complete each of the reproductions.

Time: 139. Scored when the total time is less than 140 seconds.

Short figures. Presence of one or more reproductions equaling or falling below the 10th

TABLE 1

CRITICAL LENGTHS AND HEIGHTS FOR SHORT, LONG, FLAT, AND TALL FIGURES

Figure	Short Figures		Long Figures		Flat Figures		Tall Figures	
	Initial (mm.)	Retest (mm.)	Initial (mm.)	Retest (mm.)	Initial (mm.)	Retest (mm.)	Initial (mm.)	Retest (mm.)
A	27	29	58	55	15	16	32	31
1	69	70	150	146	—	—	—	—
2	86	82	171	168	13	13	26	25
3	25	25	53	54	22	22	46	47
4	33	31	55	52	35	32	57	54
5	30	31	57	56	32	31	60	58
6	86	82	156	147	55	52	96	92
7	30	30	55	53	39	36	68	63
8	63	58	108	104	13	12	24	21

TABLE 2

CRITICAL RATIOS OF SIGNS SIGNIFICANT AT THE 5 PER CENT LEVEL ON EITHER THE INITIAL TEST OR THE RETEST

Sign	Critical Ratio		Sign	Critical Ratio	
	Initial	Retest		Initial	Retest
Global			Individual		
Second sheet	−2.19	Insig.	80-4	Insig.	+2.07
Short figures	−3.21	−3.85	14-5	+2.30	Insig.
Cohesion: ½	−2.22	−2.29	17-5	Insig.	−2.42
Time: 139	+3.00	Insig.	27-5	−2.96	−2.42
Individual			29-5	−2.76	−4.62
10-A	Insig.	−3.34	31-5	Insig.	−2.65
27-A	−3.25	Insig.	49-5	−2.39	Insig.
30-A	Insig.	−2.62	57-5	−2.11	−2.29
45-A	−2.06	Insig.	59-5	Insig.	−2.44
75-A	Insig.	−2.66	60-5	−3.71	−2.78
16-1	−2.32	Insig.	20-6	−5.73	−2.50
20-1	Insig.	−2.03	22-6	−2.73	−4.04
36-1	−2.29	Insig.	23-6	Insig.	−2.28
55-1	−2.67	Insig.	29-6	−4.99	−5.43
60-1	−5.58	−2.05	31-6	+2.67	Insig.
61-1	−2.70	−2.39	47-6	Insig.	−2.11
70-1	+2.38	+2.63	58-6	+2.57	+2.06
37-2	+2.26	Insig.	60-6	−3.30	−2.93
55-2	Insig.	−3.76	61-6	−2.26	−2.82
60-2	−3.29	Insig.	5-7	Insig.	−1.99
61-2	−3.01	Insig.	10-7	−2.31	Insig.
64-2	−2.20	Insig.	27-7	−2.30	Insig.
15-3	Insig.	−2.14	30-7	Insig.	−2.06
32-3	+3.11	Insig.	65-7	+2.52	Insig.
60-3	−6.65	−3.13	66-7	−3.49	−4.71
61-3	−4.03	−2.39	68-7	Insig.	+2.20
29-4	−2.78	Insig.	72-7	Insig.	−2.29
31-4	−2.05	Insig.	27-8	−3.99	Insig.
44-4	Insig.	−2.18	30-8	Insig.	−2.07
57-4	+7.52	+7.74	33-8	Insig.	−2.32
59-4	+3.67	+2.99	40-8	Insig.	−2.63
72-4	Insig.	−3.28	72-8	−2.05	Insig.
75-4	−4.12	−3.36	74-8	−2.76	Insig.
			75-8	Insig.	−2.49

Note: A minus sign before the critical ratio indicates that the scoring sign occurs more frequently in the total neurotic group ($N=108$); a plus sign indicates a greater incidence in the total control group ($N=285$).

percentile of the total control group in length. For the initial test and the retest, the critical lengths are given in Table 1.

Long figures. Presence of one or more reproductions equaling or exceeding the 90th percentile of the total control group in length. For the initial test and the retest, the critical lengths are given in Table 1.

Flat figures. Presence of one or more reproductions equaling or falling below the 10th percentile of the total control group in height. For the initial test and the retest, the critical heights are given in Table 1.

Tall figures. Presence of one or more reproductions equaling or exceeding the 90th percentile of the total control group in height. The critical heights for the initial test and the retest are given in Table 1.

IV. RESULTS

A. Global and Individual Figure Signs on the Initial Test and the Retest

As a first step in the analysis of the Bender-Gestalt Test records, the percentage incidence of each scoring sign was found separately for the total control and total neurotic groups. Critical ratios of the differences between each pair of percentages were then computed. Table 2 presents those global and individual figure signs yielding critical ratios significant at the 5 per cent level or better on either the initial test or the retest. Only four of the twelve global signs (second sheet, short figures, cohesion: ½, and time: 139) are significant at the 5 per cent level on the initial test, and this number is reduced to two (short figures, cohesion: ½) on the retest. The order of placement of the reproductions on the page (connections, collision, rigid sequence), numbering the reproductions, placing the reproductions in compartments, variations in height, and excessive length are nondiscriminating variables. Most consistently significant are spatial compression and shortness of length of the reproductions. Of the individual figure signs, 40 are significant on the initial test and 41 on the retest.

In order to develop a scoring system which would be intrinsically stable and reliable rather than dependent upon the summation of chance differences, a sign was required to satisfy either of the following criteria before being selected for inclusion in the final scoring scheme: (*a*) discrimination between the total control and total neurotic groups at the 5 per cent level (critical ratio of 1.96) on either the initial test or the retest, and at the 10 per cent level (critical ratio of 1.65) on the other, or (*b*) discrimination at the 1 per cent level (critical ratio of 2.58) on either test and consistency in the direction of the difference on the other. Thirty-one individual scoring signs and three global signs satisfy these criteria.

Since, however, the two major criterion groups differ somewhat with respect to age composition, educational background, and marital status, even though they are essentially alike with regard to nationality, race, sex, and military service, the possibility cannot be ignored that differences in these three variables may conceivably account for some of the obtained differences. In order to give due consideration to this possibility, an attempt was made to match as many of the neurotic subjects as possible with an equal number of subjects from the below-70 group with respect to the three variables under discussion. The below-70 group was selected as the contrasting group because it satisfied both the clinical and psychometric criteria of normality adopted for this investigation. A total of 138 subjects was so matched—69 neurotics and 69 controls—and will be referred to hereafter as the neurotic matched criterion group and the below-70 matched criterion group. While perfect matching could not be achieved because of the requirement that each case be matched with respect to

TABLE 3

AGE, GRADE, AND MARITAL STATUS COMPOSITION OF THE MATCHED CRITERION GROUPS

	Below 70 ($N=69$)	Neurotic ($N=69$)
Age (years)		
Mean	27.26	27.84
SD	4.70	4.34
Education (grades)		
Mean	12.19	12.19
SD	2.05	2.05
Marital Status		
Single	37	32
Married	32	33
Divorced	0	1
Separated	0	3

three variables simultaneously, the success of the matching technique is evident from Table 3.

The percentage incidence of the 34 tentatively selected scoring signs was found separately for the normal and neurotic matched groups. Critical ratios of the differences between each pair of percentages were then computed. Since the number of cases is smaller in the matched groups than in the original groups, the critical ratios tend to be smaller even when the absolute differences are identical. As a basis for inclusion of the signs in the final scoring system, therefore, the criterion set for this second weeding-out process was somewhat more lenient, namely, discrimination at the 5 per cent level on either the initial test or the retest and consistency in the direction of the difference on the other. Table 4 presents the initial and retest critical ratios of the 27 individual figure signs and the three global signs which survived both screening procedures.

The 30 signs which constitute the final scoring system may be grouped as follows:

Description	*Sign*[3]
Numeration (number of wave crests in figure 6)	20–6
Wave flattening in figure 6	22–6
Asymmetry in figure 5	27–5
Displacement in figures 5 and 6	29–5 29–6
Contiguity in figure A	30-A
Parallel lines in figure 8	33–8

TABLE 4

CRITICAL RATIOS OF THE SELECTED SIGNS FOR THE MATCHED CRITERION GROUPS

Sign	Critical Ratio	
	Initial	Retest
Short figures	−2.46	−2.81
Cohesion: ½	−2.58	−2.68
Time: 139	+1.96	+0.34
20-6	−2.79	−0.35
22-6	−0.65	−3.00
27-5	−2.56	−0.81
29-5	−2.02	−1.19
29-6	−0.87	−1.98
30-A	−0.40	−3.59
33-8	−1.96	−1.88
55-1	−2.05	0.00
55-2	−1.66	−2.77
57-4	+4.02	+3.59
70-1	+0.75	+2.28
72-4	−0.85	−3.24
57-5	−2.32	−2.06
58-6	+2.77	+1.31
60-1	−2.44	−2.67
60-2	−3.20	−1.90
60-3	−4.76	−2.30
60-5	−4.73	−3.62
60-6	−2.92	−3.22
61-1	−2.18	−2.25
61-2	−2.63	−1.93
61-3	−2.97	−1.83
61-6	−1.75	−3.01
65-7	+3.50	+2.45
66-7	−3.20	−4.33
75-A	−0.68	−2.51
75-4	−3.01	−2.50

Note: A minus sign before the critical ratio indicates that the scoring sign occurs more frequently in the neurotic group ($N=69$); a plus sign indicates a greater incidence in the below-70 group ($N=69$).

---

[3] Signs followed by a plus occur more frequently among normals than among neurotics; all other signs discriminate in the opposite direction, i.e., are more characteristic of neurotics.

Description	*Sign*
Upward slope in figures 1 and 2	55–1
	55–2
Sinistrad direction in figures 4 and 5	57–4+
	57–5
Counting of test figure in figures 1, 2, 3, 5, and 6	60–1
	60–2
	60–3
	60–5
	60–6
Upward direction in figure 6	58–6+
Counting of reproduction in figures 1, 2, 3, and 6	61–1
	61–2
	61–3
	61–6
Initial part (left hexagon made first) in figure 7	65–7+
Angle differentiation in figure 7	66–7
Pairing of dots in figure 1	70–1+
Correction in figure 4	72–4
Sagging line in figures A and 4	75-A
	75–4
Spatial compression	Cohesion: ½
One or more reproductions of abbreviated length	Short figures
Total time less than 140 seconds	Time: 139+

B. The Recall Test

The mean number of figures recalled by the total control group is 6.28, compared with a mean of 5.93 for the total neurotic group. The standard deviations are 1.15 and 1.60, respectively, and the critical ratio of the difference between means is only 1.79. There is, however, a tendency for the neurotic group to contain more instances of extreme paucity of recall. Thus 12.0 per cent of the neurotics recall 3½ figures or less compared with 1.8 per cent of the controls. If one focuses on the frequency of recall of each of the figures rather than on the total number of figures recalled, only figures A and 6 show differences in percentage incidence significant at the 5 per cent level (critical ratios of 2.29 and 2.24, respectively).

Inspection of a number of recall test records suggested that when the figures were reproduced from memory, certain modifications were present which either infrequently or never occurred during the initial and final tests (when the reproductions were copied from models). These changes were classified as distortions, confusions, half figures, closures, and rotations. Distortions occur very frequently regardless of the group to which the subject belongs, 57 per cent of the below-70 matched criterion group and 64 per cent of the neurotic matched criterion group having one or more distorted reproductions in their recall records. The differential incidence of all special modifications is, however, statistically insignificant, except for closure on figure 4.

Only six signs meet the criterion set for the recall test, namely, discrimination between the total control and total neurotic groups at the 5 per cent level. These signs and their critical ratios (of the differences between percentages) are shown in Table 5. Closure-4 refers to reproducing the open square in figure 4 as a closed square. A (5-9) refers to recall of figure A as the fifth, sixth, seventh, eighth, or ninth figure recalled by the subject, while 6 (1-4) refers to recall of figure 6 among the first four figures recalled. Short figures-4 refers to the presence of four or more reproductions at or below the 10th percentile of the initial

TABLE 5

Critical Ratios of Signs Included in the Final Scoring System for the Recall Test

Sign	Critical Ratio
Cohesion: ½	−2.13
Closure–4	+2.21
Short figures–4	−2.69
A (5–9)	+3.58
6 (1–4)	+4.24
3½ or less	−3.16

Note: A minus sign before the critical ratio indicates that the scoring sign occurs more frequently in the total neurotic group ($N = 108$); a plus sign indicates a greater incidence in the total control group ($N = 285$).

TABLE 6

Means and Critical Ratios of the Differences between Means of the Initial, Retest, and Combined Scores

Group	*N*	Initial Score		Retest Score		Combined Score	
		Mean	CR	Mean	CR	Mean	CR
Below 70 (matched) Neurotic (matched)	69 69	3.07 8.06	9.78	2.10 6.46	8.90	5.17 14.52	11.00
Below 70 (unmatched) Neurotic (unmatched)	86 39	3.56 8.59	7.62	2.55 6.77	7.96	6.12 15.36	8.72
Below 70 1 above 70	155 68	3.34 3.82	1.20	2.35 2.75	1.14	5.70 6.57	1.32
Below 70 2 above 70	155 62	3.34 3.89	1.38	2.35 3.23	2.44	5.70 7.11	2.07
1 above 70 2 above 70	68 62	3.82 3.89	.15	2.75 3.23	1.12	6.57 7.11	.66
Total control Total neurotic	285 108	3.58 8.26	13.00	2.63 6.59	12.00	6.26 14.85	14.08

test reproductions of the total control group in length. The recall of less than four figures by the subject is designated by the sign, 3½ or less.

C. Scoring of the Test Records of the Criterion Groups with the Selected Signs

Following the isolation of those signs on the initial test, recall test, and retest which individually discriminated between the normal and neurotic subjects, the next step was to rescore all the records for these "selected signs," add the signs algebraically, and tabulate the distributions of these summated scores. A combined score, defined as the sum of the initial and retest scores of each subject, was also obtained. The presence of a sign which occurs more frequently among neurotic subjects is indicated by a minus sign, and the presence of a sign occurring more often among the controls is designated by a plus. The sums of these signs appear without sign when the number of minus signs outweighs the plus signs, and appear with a plus sign when the number of plus signs outweighs the negative signs. If, for example, a given test record has six negative signs and three positive signs, the sum is designated as 3; should the record contain four positive signs and three negative signs, however, the sum would be designated as + 1. The results of this procedure for the matched and unmatched criterion groups, the three control subgroups, and the total control and total neurotic groups are presented in Table 6, which gives the means and critical ratios of the difference between means of the initial, retest, and combined scores.

All the differences between the control and neurotic groups are highly significant, and overlapping of the distributions of scores is not great. In the case of the below-70 and neurotic matched groups, for example, an initial test score of 5 screens out 88.4 per cent of the neurotics at the expense of only 26.1 per cent false positives; a retest score of 4 screens out 81.2 per cent of the neurotics with only 23.2 per cent false positives; and a combined score of 8 eliminates 89.0

per cent of the neurotics at a cost of 23.2 per cent of the controls. None of the differences between the three control subgroups is significant at the 1 per cent level, although the trend of the means is in accord with expectation. With regard to the recall scores, the means of the below-70 and neurotic matched groups are +.89 and .07, respectively, a difference yielding a critical ratio of 4.82. This difference is much smaller than that obtained for the initial, retest, and combined scores. Various combinations of the recall score with each of these scores was tried, but the general finding was that the recall score did not sufficiently increase the discriminating power of the test to warrant its inclusion in a more complex composite score.

D. EFFECT OF EDUCATION, AGE, AND INTELLIGENCE UPON TEST SCORES

The similarity of the results for matched and unmatched groups suggests that such factors as education and age bear little or no relationship to the Bender-Gestalt Test scores of adult subjects. One would therefore expect that intelligence would be uncorrelated with Bender-Gestalt Test performance. These expectations are confirmed by Table 7, which summarizes the correlations of these three variables with the initial and retest scores of the total control and total neurotic groups.

E. SCORING OF THE INITIAL TEST RECORDS OF THE MATCHED CRITERION GROUPS WITH PASCAL AND SUTTELL'S SIGNS

Pascal and Suttell (35) are the only investigators other than the present writer who have combined Bender-Gestalt signs into a total score which, it is claimed, will differentiate normals from abnormals with a reasonably high degree of accuracy. Pascal and Suttell claim that their scores will distinguish normals from neurotics, normals from psychotics, and neurotics from psychotics, although the latter differentiation is admittedly more subject to error. From the standpoint of the subject of the present paper, the main interest centers around the relative effectiveness of Pascal and Suttell's signs in distinguishing between the normal and neurotic cases employed in the present investigation. This question assumes significance in view of the fact that many of the signs used by Pascal and his collaborator were identical or similar to signs tentatively tried out in the present study but rejected for the final scoring system because of lack of statistical validity. The below-70 and neurotic matched criterion groups were selected for this check because they are sharply contrasting groups and because such factors as race, nationality, sex, age, education, and marital status were held constant, eliminated, or neutralized.

Accordingly, the present writer learned Pascal and Suttell's scoring scheme and

TABLE 7

CORRELATIONS OF GRADE, AGE, AND INTELLIGENCE TEST SCORES WITH INITIAL AND RETEST SCORES

	Total Control			Total Neurotic		
	N	Initial	Retest	*N*	Initial	Retest
Grade	282	−.11 ± .04	−.10 ± .04	108	−.12 ± .06	−.06 ± .07
Age	277	.06 ± .04	.04 ± .04	107	.20 ± .06	.16 ± .06
Otis IQ	285	−.15 ± .04	−.17 ± .04			
W-B IQ				84	−.10 ± .07	−.15 ± .07

TABLE 8

Raw Score Means and Critical Ratios of the Differences between Means on the Initial Test for the Matched Groups, Scored According to Pascal and Suttell's System

Grade	Mean		Critical Ratio
	Below 70 ($N=69$)	Neurotic ($N=69$)	
8–11	25.15	28.60	.72
12	24.20	27.55	.99
College	27.45	29.40	.62
Total group	25.55	28.45	1.38

scored several of the test records which appear in their book until his total scores closely approximated those given in their scoring manual. All 138 cases in the matched groups were then scored in accord with this method. The raw score means and critical ratios of the differences between means are given in Table 8.

Education is unrelated to the size of the score of either the controls or the neurotics, as is evident from the fact that subjects who did not receive high school diplomas, high school graduates, and college trained subjects obtain almost identical mean scores within the criterion group of which they are members. Pascal and Suttell claim that education is a significant factor and present separate normative data for college and non-college trained individuals. Of much greater significance, however, is the failure of their scoring scheme to distinguish between any of the normal and neurotic groupings.

Pascal and Suttell's administration of the Bender-Gestalt differs from that employed in the present study in two respects as far as the initial test is concerned: (*a*) the time required to reproduce the figures is not recorded, and (*b*) sketching is prohibited. However, sketching occurs relatively infrequently in the matched group cases, and it is difficult to see how the single factor of recording time could wipe out the presumed differences in test performance which are implicit in their method of scoring.

V. THE CROSS-VALIDATION STUDY

If Pascal and Suttell's scoring scheme holds up so poorly under the strain of cross validation, one may well speculate as to the fate of the scoring method developed in the present investigation under similar circumstances. The procedure of scoring the test records of the very subjects upon whom the item analysis is based is open to the objection that the summation of chance differences may result in the spuriously high differential that cannot subsequently be duplicated when applied to other samples drawn from the same populations.

Ideally, it would have been desirable to have obtained samples of normal and neurotic veterans drawn from the same populations upon which the item analysis was performed, but this procedure was not feasible because of the virtually insurmountable difficulty of obtaining additional neurotic cases from the Veterans Administration mental health clinics in New York City during a period of peak load. The thought then occurred that the cross-validation study might profitably be conducted with nonveteran male adults differing markedly in personal adjustment, for if the test held up under these conditions the conclusions concerning validity would not have to be restricted to the veteran population but would have a wider applicability to the male population at large.

It was decided to select the cases from the files of the New York University Testing and Advisement Center, a testing and vocational guidance agency that serves the metropolitan area and suburbs. The vast majority of clients 18 years of age and over who avail themselves of the Center's services desire educational or vocational guidance. However, a certain proportion of these individuals are subsequently found to have personality problems sufficiently severe to warrant the recommendation that they undergo psychotherapy as an initial step in the solution of their vocational problems. The testing and counseling procedures consist of (*a*) an initial interview, in which personal history data regarding education, employment experience, health, family background, marital adjustment, interests, hobbies, special talents, and personal and social adjustment are obtained; (*b*) 10 to 16 hours of group and individual testing, including the Wechsler-Bellevue, educational achievement tests, aptitude tests, interest tests, personality inventories, and projective personality tests; (*c*) one or more counseling conferences in which the findings are presented to the client and plans for the future are mutually worked out; and (*d*) a detailed written report which integrates biographical and test data and recapitulates the major findings. Since comparatively few of these clients had educations below the high school level,

TABLE 9

COMPOSITION OF THE CROSS-VALIDATION GROUPS

Group	*N*	Education		Age		Marital Status		
		Mean (grade)	*SD*	Mean (years)	*SD*	% Married	% Single	% Other
TAC Male								
Satisfactory adjustment	29	14.41	1.67	25.56	8.04	31.0	69.0	0.0
Unsatisfactory adjustment	29	14.41	1.67	25.68	7.60	20.7	79.3	0.0
TAC Total Male								
Satisfactory adjustment	39	14.36	1.48	25.42	7.34	33.3	66.7	0.0
Unsatisfactory adjustment	43	13.98	1.71	27.56	8.42	20.9	74.4	4.7
TAC Total Female								
Satisfactory adjustment	15	13.07	1.34	26.60	8.36	20.0	73.3	6.7
Unsatisfactory adjustment	21	13.71	1.64	33.10	8.82	33.3	61.9	4.8
TAC Total								
Satisfactory adjustment	54	14.00	1.55	25.74	7.66	29.6	68.5	1.9
Unsatisfactory adjustment	64	13.89	1.69	29.38	8.94	25.0	70.3	4.7

the cross-validation cases were selected from a file of approximately one thousand clients ranging in age from 18 to 50 and in education from four years of high school through college. Each of these cases was evaluated from the standpoint of whether the subject was effecting an unusually good or an exceptionally poor emotional and social adjustment. Only when personal history data, Wechsler-Bellevue scatters, and personality tests pointed consistently in one direction or the other and only when these data were consistent with the impressions of the psychologist who originally handled the case was the subject selected for inclusion in the cross-validation groups. The Bender-Gestalt Test records were removed from the folders prior to this appraisal so that they would not influence the final evaluation. The Bender-Gestalt Tests were in all instances administered with the same instructions that had previously been used for the veteran subjects but they had not been scored, since no scoring scheme was available at the time. Most of the poorly adjusted subjects who were selected would probably have been diagnosed as neurotic by a psychiatrist (no cases in which there was any suspicion of psychosis were included), but the question of diagnosis was avoided by merely designating the cases as satisfactorily or unsatisfactorily adjusted.

A total of 118 cases which satisfied the above criteria was finally selected. Of these 54 (39 males and 15 females) were designated as the TAC total satisfactory adjustment group, and 64 (43 males and 21 females) as the TAC total unsatisfactory adjustment group. ("TAC" refers to "Testing and Advisement Center.") Originally, it had been intended that only males would be included, but a smaller number of female records was added in order to shed some light on the question of whether the method of scoring is valid for both sexes. These groups were subdivided into several subgroups which are described in Table 9 with regard to size, education, age, marital status, and sex. As a first step, the male subjects were matched, pair by pair, for educational background, age, and marital status. Matching was possible for 58 of the 82 male cases, and the two groups so matched were designated as the TAC male satisfactory adjustment group ($N = 29$) and the TAC male unsatisfactory adjustment group ($N = 29$). The "TAC total male satisfactory adjustment" and "TAC total male unsatisfactory adjustment" groups refer to all the males in the cross-validation study; the "TAC total female satisfactory adjustment" and "TAC total female unsatisfactory adjustment" groups refer to all the females.

Tables 10 and 11 recapitulate the findings in the cross-validation study with respect to the initial, retest, and combined scores. Table 10 may be compared with Table 6, which presents analogous results for the original veteran groups. Shrinkage in the size of the difference between means (and hence in validity) is evident, but discrimination is still relatively good. For the entire group of 118 cases, the initial and combined scores are roughly equivalent in discriminating power, and both are slightly superior to the retest. An initial score of 6 or above and a combined score of 10 or above each screen out 70.3 per cent of the poorly adjusted cases at the cost of 40.7 per cent false positives; a retest score of 4 or above, on the other hand, has a false positive rate of 51.9 per cent for a pick-up of 71.9 per cent. With regard to the females, the initial score means of 4.60 and 7.14 and the combined score means of 6.93 and 11.67 yield statistically significant

TABLE 10

MEANS AND CRITICAL RATIOS OF THE DIFFERENCES BETWEEN MEANS OF THE INITIAL, RETEST, AND COMBINED SCORES FOR THE CROSS-VALIDATION GROUPS

Group	N	Initial Score		Retest Score		Combined Score	
		Mean	CR	Mean	CR	Mean	CR
TAC Male							
Satisfactory adjustment	29	5.45	2.96	4.45	3.66	9.90	3.64
Unsatisfactory adjustment	29	7.73		7.34		15.07	
TAC Total Male							
Satisfactory adjustment	39	5.13	3.80	3.92	4.15	9.05	4.53
Unsatisfactory adjustment	43	7.44		6.74		14.30	
TAC Total Female							
Satisfactory adjustment	15	4.60	2.82	3.27	1.25	6.93	2.56
Unsatisfactory adjustment	21	7.14		4.52		11.67	
TAC Total							
Satisfactory adjustment	54	4.98	4.98	3.74	4.03	8.46	5.08
Unsatisfactory adjustment	64	7.42		6.02		13.44	

differences, but the retest score means of 3.27 and 4.52 do not, although the difference is in the expected direction. In view of the fact that two of the three kinds of scores shown to be significant for the males are also significant for the females, it is probable that the scoring method developed in this study is applicable to either sex and that the lack of completely parallel findings is attributable to the small size of the female samples. As expected from its comparatively poor showing with the original veteran groups, the recall test fails to hold up under cross validation. The means of the TAC total satisfactory adjustment and TAC total unsatisfactory adjustment groups are + .74 and + .62, respectively, and the critical ratio of the difference is only 0.64.

VI. DISCUSSION

Perhaps the most significant finding of the present investigation is that normals and neurotics are much more alike than different with regard to perceptual-motor response to simple geometric designs, to the extent that such responses are measured by the reproduction of the Bender-Gestalt figures from a model or from memory. Only about one-sixth of the deviations for which the reproductions were scored discriminate between the normal and neurotic subjects in the initial validation study, even when the comparatively lenient criterion of differ-

TABLE 11

CUMULATIVE PERCENTAGE DISTRIBUTIONS OF SCORES FOR ALL CASES IN THE CROSS-VALIDATION STUDY*

Score	Initial Test		Retest		Combined Score	
	Sat. cf%	Unsat. cf%	Sat. cf%	Unsat. cf%	Sat. cf%	Unsat. cf%
25						100.0
24						98.4
23						96.9
22					100.0	96.9
21					98.1	93.8
20					98.1	92.1
19					98.1	84.4
18					96.3	78.1
17					96.3	71.9
16		100.0		100.0	96.3	64.1
15		98.4		98.4	90.7	59.4
14		98.4		98.4	90.7	53.1
13		98.4		98.4	88.9	50.0
12		98.4		96.9	85.2	42.2
11	100.0	93.8	100.0	95.3	77.8	39.1
10	96.3	85.9	98.1	89.1	68.5	35.9
9	96.3	81.3	98.1	81.3	59.3	29.7
8	90.7	57.8	98.1	76.6	50.0	25.0
7	85.2	46.9	94.4	60.9	38.9	23.4
6	72.2	37.5	83.3	51.6	37.0	15.6
5	59.3	29.7	79.6	43.8	27.8	9.4
4	48.1	15.6	66.7	37.5	16.7	4.7
3	31.5	10.9	48.1	28.1	13.0	1.6
2	11.1	4.7	33.3	17.2	11.1	1.6
1	7.4		14.8	14.1	7.4	
0	3.7		5.6	7.8	3.7	
+1			3.7	3.1	1.9	
+2			1.9		1.9	

* TAC total satisfactory adjustment group, $N=54$; TAC total unsatisfactory adjustment group, $N=64$.

entiation at the 5 per cent confidence level is invoked, and the number of discriminating signs shrinks considerably if a more rigorous test of statistical significance is applied. This result is in essential agreement with that reported by Billingslea (17), who found only 12 out of 63 indices (19 per cent) significant at the 5 per cent confidence level. It agrees also with Bender (12) and Woltmann (51) who, without reporting quantitative data, conclude that the Bender-Gestalt is of limited value in distinguishing normal from neurotic individuals. It is in marked disagreement with Hutt (28), who speaks of a "neurotic syndrome" on the Bender-Gestalt without, however, offering any statistical support for his conclusion, and with Pascal and Suttell (35), who report that 105 of their 200 signs (52.5 per cent) differentiate between their patient and nonpatient groups. It is only fair to point out that in the latter instance a sign is regarded as significant if it distinguishes normals from a combined neurotic and psychotic group of patients, and that their figure would be somewhat smaller if the psychotics were excluded from the comparisons.

Modifications of the stimulus figures are frequent in both neurotic and normal subjects, according to the findings of the present study, but these deviations rarely approach the point where they can be described as gross destructions of the Gestalt, except when the reproductions are drawn from memory in the absence of an objective stimulus; even under such circumstances the incidence of distortions is approximately equal for the two contrasting criterion groups. The writer would agree with Billingslea and with Pascal and Suttell that gross deviations are nondiscriminating and that, if any separation is to be effected, more subtle modifications must be defined, observed, and isolated. The writer would further agree that a feasible scoring scheme should take into account these minute deviations, but exception is taken with Billingslea's belief that such a method of scoring should be accomplished by equally precise physical measurements. Billingslea took as long as fifteen hours to score a single record—a procedure which might be justified in the research laboratory but not in the psychological clinic. The writer agrees with Pascal and Suttell that scoring should be accomplished by inspection rather than measurement but that the scoring criteria should be so carefully described that no great loss is experienced in going from quantitative measurement to qualitative observation. Moreover, this qualitative inspection should be supported by quantitative normative data rather than by loose generalizations based on clinical experience with a predominantly abnormal population.

In the present investigation, none of the "abnormal" signs occurs exclusively in the abnormal subjects, and none of the "normal" signs is found solely in the control subjects—a finding similar to that reported by Anastasi and Foley (5) in their experimental study of drawings but differing from the results of Pascal and Suttell, who give maximum weight to signs having zero frequencies in their control group and frequencies of one or two in their patient group. It might also be noted in this connection that one might have entertained any one of four hypotheses prior to the present study: (*a*) deviations of the reproductions from the stimulus figures are more frequent among abnormals than normals, i.e., normals tend to conform more strictly to the stimuli; (*b*) deviations are more frequent among normals, i.e., abnormals tend to conform more rigidly to the stimuli; (*c*) deviations are equally characteristic of both normals and abnormals; and (*d*) certain deviations are more characteristic of normals and others more characteristic of abnormals. Most clinicians implicitly assume hypothesis *a*, a corollary of which is that the more nearly the reproductions approximate the originals, the more likely the examinee is to be normal; and that the more the record deviates from the original the greater the severity of the disturbance. The hypothesis that deviations are more characteristic of abnormals is supported by the fact that in the present investigation 31 of the 40 initial test signs (77.5 per cent) and 35 of the 41 retest signs (85.4 per cent) which discriminate at the 5 per cent level are negative in sign, i.e., occur more frequently among neurotics. Hypotheses *b* and *c* are rejected on this basis, while hypothesis *d* is given support. Billingslea also found signs which discriminated in both directions, and such signs are implicit in Hutt's formulations. The two relevant hypotheses may therefore be combined and restated as a conclusion, namely, that deviations of the reproductions from the originals occur more often in the records of maladjusted individuals, and that both types of deviations should be considered in evaluating Bender-Gestalt Test performance. Whether the degree of psychological disturbance progressively parallels an increase in the excess of "abnormal" over "normal" signs cannot be ascertained from the present study, since the subjects were not graded according to severity of disturbance and since, unlike Pascal and Suttell's study, psychotic subjects were not tested. A high correlation between size of score and severity of disturbance seems unlikely, however, in view of the high degree of overlapping of the control and experimental groups, an overlapping which is especially pronounced in the cross-validation study. Cut-off scores may be set which will separate normals and abnormals in varying proportions, the normals tending to make more favorable (lower) scores and the abnormals more unfavorable (higher) scores, but as the critical score is raised there is a steady increase in false negatives. Although it is likely that psychotics will tend to make higher scores than neurotics and normals, the interpretation of increasing severity of disturbance with increasing score within the neurotic or maladjusted groups in the present investigation is not warranted by the data. The presence of three or four signs which occur very rarely in the standardization group (e.g., total or part rotation, dissociation, fragmentation, compartments, numbering, perseveration, confused sequence) may be much more significant than the occurrence of eight or nine deviations that are found more often in both groups. It should also be noted that the scoring scheme developed in the present study is intended to differentiate adult normals from adult neurotics and poorly adjusted individuals and that its applicability to psychotics, organics, children, mental defectives, etc. is not known. There is no reason, however, why the method could not be extended to such nosological groups by appropriate item-analysis techniques.

The hypothesis that differences in test performance between normals and neurotics would be accentuated in the absence of an objective stimulus during recall is not given strong support by our findings. Only six discriminating signs could be isolated in the initial validation study, and the difference between the means of the criterion groups was considerably less than those obtained for initial and retest scores. When applied to the cross-validation groups, the differences on the recall test disappeared almost completely. If, however, a retest or a combined score is desired, it is necessary to administer the recall test, since the original data were obtained under these conditions. It seems likely that the retest and combined scores would not differ markedly from those obtained in the present investigation, even if the recall test were omitted, but this is a problem for future research, and until the results of such research are forthcoming, it is recommended that the present method of administration be adhered to. If time is limited, the difficulty may be surmounted by giving only the initial test, for the results obtained under these conditions are in essential agreement with those found with the longer procedure, except for the fact that greater confidence can be attached to the findings when initial, retest, and combined scores are consistent in their diagnostic implications.

With regard to the scoring method developed in the present study, the prediction that it would be successful in differentiating well-adjusted from poorly adjusted individuals is verified by the results of the cross-validation study. The differences between the means of these contrasting groups are smaller than those obtained in the initial validation, but the differences emerge as statistically significant despite the anticipated shrinkage in validity. The fact that the scoring scheme can be applied to nonveterans as well as veterans, to females as well as males, and to individuals who are maladjusted but not necessarily neurotic, and the additional fact that differences in age (within the range of 18 to 50), education (within the high school and college levels), and intelligence (for IQ's of 85 and above) do not appreciably influence the score indicates that the Bender-Gestalt can serve a useful screening function, not only in mental hygiene clinics but also in vocational guidance and psychological counseling agencies dealing with normal individuals and persons representing varying degrees of emotional disturbance short of psychosis. It is recommended as a screening device rather than as an instrument for elaborate personality interpretation because the differential incidence of most of the deviations is not great and because many of the signs said to have particularized meaning for neurotics prove to be nondiscriminating. Interpretations of "flexibility" or "spontaneity" based on logical sequence, "deterioration" or "emotional immaturity" based on substitution of dashes or circles for dots, "depression" based on downward slope, "elation" based on upward slope, "instability" based on curvature irregularities, "neurotic sex conflict" based on closure or elongation or repetitive sketching of projections, "inferiority" or "insecurity" based on reduction in size, and "expansiveness" or "oppositional traits" based on increase in size appear to be premature in the light of the available evidence. Clinicians working almost exclusively with abnormal patients may establish certain correlations between some of these signs and information gathered about the patient from behavioral, clinical, or psychometric data, but it is submitted that they might well experience a keen feeling of disillusionment at the absence of such correlations in work with comparatively well-adjusted people.

The cut-off score to be used for screening purposes will depend upon the particular purpose to be achieved. If it is desired to screen out a great proportion of the maladjusted individuals in a given sample, the critical score must be lowered despite the relatively high incidence of false positives. Reference to the distributions of initial, retest, and combined scores in Table 11 should prove helpful to the clinician in interpreting a given score in terms of the chances that such a score is likely to be made by an emotionally disturbed individual. These tables may readily be converted into expectancy tables in personnel-selection situations where there is greater interest in group prediction than in individual predictions. For general purposes, it is suggested that cut-off scores be set at one standard deviation above the mean of the control groups in Table 11. Disregarding fractions, these critical scores are 7, 6, and 13 for the initial, retest, and combined scores, respectively. On the initial test, a score above 7 screens out 53.1 per cent of the maladjusted cases at the cost of 14.8 per cent false positives; on the retest, a score above 6 screens out 48.4 per cent at the cost of 16.7 per cent false positives; a combined score above 13 picks up 50.0 per cent of the maladjusted cases at the expense of 11.1 per cent of the normals.

What can be said of the reliability of the scoring system developed in the present study? The concept of reliability is inherent in the method used to determine item validity, since only those signs which differentiated in a consistent direction on both the initial test and the retest were retained for inclusion in the final scoring system. Correlation of the initial and retest scores for the entire initial validation group of 393 cases resulted in a reliability coefficient of .67, while a similar correlation for the 82 male subjects in the cross-validation study gave rise to a coefficient of .68. These figures are remarkably close to the test-retest coefficient of .71 reported by Pascal and Suttell for 44 subjects after an interval of 24 hours. The present writer takes the position that Bender-Gestalt test-retest correlation coefficients are not true reliability coefficients, but are underestimations of the true reliability. As is true on most performance tests—and the Bender-Gestalt is no exception—a test administered a second time does not measure the same factors tapped by the initial administration. In the case of the Bender-Gestalt, the initial test constitutes an unfamiliar task, while the retest represents a relatively familiar situation. Each succeeding design can be anticipated before it is actually presented, and problems of size and spatial arrangement can be readily solved on the basis of prior experience. Memory factors, such as remembering the number of dots in figure 1 or the number of columns in figure 2, may obviate counting of the elements in the stimulus figures or reduce the time required to reproduce the figures. If the interval between test and retest is excessively long, these memory factors will be reduced to a minimum, but changes in the subject's adjustment or motivation may preclude consistency of measurement. Analysis of the present findings shows that there is better organization and apportionment of space on the retest, resulting in more logical sequence, that less time is required to reproduce the figures, that the reproductions tend to be smaller, and that counting is less frequent on the second administration of the test. It may be inferred from the above, although it cannot be demonstrated, that the subjects approach the retest with greater confidence and that the higher degree of structuration of the retest makes it a more pleasurable or less trying experience. The extent to which the intervening recall test influences the test-retest correlation cannot be determined, but it probably is not great in view of the similarity of the present correlations with those reported by Pascal and Suttell, who did not employ the recall situation. It is true, however, that our retest scores are systematically lower than initial test scores, whereas Pascal and Suttell find a negligible practice effect. However, they did not score for such factors as counting and time, both of which contribute markedly to a decreased retest score.

VII. SUMMARY AND CONCLUSIONS

The Bender-Gestalt Test was administered to 393 white male adult veterans (108 neurotics and 285 controls) in an initial validation study and to 118 white nonveteran adults of both sexes (64 poorly adjusted individuals and 54 controls) in a cross-validation study. The test was administered twice to each subject in conjunction with an interpolated test of immediate recall. A total of 1,533 test records (511 initial tests, 511 retests, and 511 recall tests) was scored and analyzed. An objective scoring system consisting of 82 general categories and 312 specific signs was developed, and the incidence of each of these signs was determined for each test record of the cases in the initial validation study. Thirty signs meeting specified criteria of statistical significance and consistency were isolated and constitute the final scoring system for the initial test and the retest. An additional six signs were evolved for the scoring of the recall test records. The original test records were then rescored with the selected signs and satisfactory separations of the normal and neurotic cases were achieved. The test records of the subjects in the cross-validation study were next scored with the same signs, and statistically significant separations of the poorly adjusted and well-adjusted cases obtained.

The results warrant the following conclusions:

1. Normals are much more alike than different with regard to perceptual-motor response to simple geometric designs, to the extent that such responses are measured by reproduction of the Bender-Gestalt figures from a model or from memory. In this respect, the findings agree to a greater extent with those of Bender and Billingslea than with those of Hutt and Pascal and Suttell.

2. Gross destructions of the Gestalt are quite rare in both normals and neurotics, except when the reproductions are drawn from memory, and even under such circumstances the incidence of distortions is approximately the same for the two contrasting groups. If any separation of normals from neurotics is to be effected, the evidence suggests that it must be achieved through consideration of more subtle modifications of the test figures.

3. Deviations of the reproductions from the stimuli occur more often in the records of the abnormal subjects, but certain deviations are more characteristic of well-adjusted individuals, suggesting that both types should be considered in evaluating test performance.

4. None of the "abnormal" signs occurs exclusively in the abnormal subjects, and none of the "normal" signs is found solely in the controls.

5. General factors said to apply to all or almost all of the figures, such as reduction in size, rotation, sketching, or angulation, are largely nonexistent, and any interpretation of test performance which assumes such factors is likely to be extravagant or misguided. The evidence supports the view that deviations are, for the most part, specific to specific figures.

6. Constituting the final scoring system evolved in this study are 30 signs which discriminate consistently between normals and neurotics on the initial test and the retest.

7. Neurotics and normals can be distinguished on the basis of initial, retest, recall, or combined initial and retest scores, but the recall score is much less discriminating than the others and fails completely on cross validation. With regard to discriminating power for all the populations sampled, the combined score is somewhat superior to the initial score, which is in turn slightly more effective than the retest score.

8. The expected shrinkage in validity is found in the cross-validation study, resulting in a greater degree of overlapping of the distributions of the maladjusted and well-adjusted cases, but differences between the means of the two contrasting groups, significant at the 1 per cent level, are nevertheless obtained for the initial, retest, and combined scores.

9. Retest reliability (initial vs. retest scores) is represented by correlations of .67 and .68 for the original and cross-validation groups, but these coefficients are regarded as underestimations of the true reliability on the grounds that a

second administration of the test does not duplicate the first administration. The chief reason for this lack of replication is that the retest is less amorphously structured for the subject because of increased familiarity with the requirements of the test situation. Though not demonstrated, it is highly probable that the reliability of the combined score is higher than that of the initial test or retest alone.

10. The scoring method developed in this investigation is not significantly related to sex, age (within the range of 18 to 50), education (for persons with eighth-grade schooling or better), or intelligence (for IQ's of 85 and above).

11. Hutt's "neurotic syndrome" is in general not substantiated by the present findings. Many of Hutt's psychodynamic interpretations are also open to question in view of the fact that a number of the deviations said to be indicative of emotional or motivational processes occur equally often among normals and neurotics.

12. The scoring method developed by Pascal and Suttell fails to differentiate the 138 normal and neurotic subjects of the matched criterion groups. Contrary to their findings, there is no significant relationship between education and test performance for adults, and this statement holds true whether Pascal and Suttell's scoring system or the one evolved in the present investigation is employed.

13. The Bender-Gestalt Test, as scored in the present study, is recommended as a screening device to be used as a supplement to other psychodiagnostic tests rather than as an instrument for the elaborate interpretation of individual personality dynamics.

REFERENCES

1. Anastasi, Anne, & Foley, J. P., Jr. A survey of the literature on artistic behavior in the abnormal. I. Historical and theoretical background. *J. gen. Psychol.*, 1941, **25**, 111-142.
2. Anastasi, Anne, & Foley, J. P., Jr. A survey of the literature on artistic behavior in the abnormal. II. Approaches and interrelationships. *Ann. N. Y. Acad. Sci.*, 1941, **42**, 1-111.
3. Anastasi, Anne, & Foley, J. P., Jr. A survey of the literature on artistic behavior in the abnormal. III. Spontaneous productions. *Psychol. Monogr.*, 1940, **52**, No. 6 (Whole No. 237).
4. Anastasi, Anne, & Foley, J. P., Jr. A survey of the literature on artistic behavior in the abnormal. IV. Experimental investigations. *J. gen. Psychol.*, 1941, **25**, 187-237.
5. Anastasi, Anne, & Foley, J. P., Jr. An experimental study of the drawing behavior of adult psychotics in comparison with that of a normal control group. *J. exp. Psychol.*, 1944, **34**, 169-194.
6. Ash, P. The reliability of psychiatric diagnosis. *J. abnorm. soc. Psychol.*, 1949, **44**, 272-276.
7. Barkley, Bill J. A note on the development of the Western Reserve hapto-kinesthetic gestalt test. *J. clin. Psychol.*, 1949, **5**, 179-180.
8. Bell, J. E. *Projective techniques: a dynamic approach to the study of the personality.* New York: Longmans, Green, 1948.
9. Bender, Lauretta. Principles of gestalt in copied form in mentally defective and schizophrenic persons. *Arch. Neurol. Psychiat.*, 1932, **28**, 661-673.
10. Bender, Lauretta. Disturbances in visuomotor gestalt function in organic brain diseases associated with sensory aphasia. *Arch. Neurol. Psychiat.*, 1933, **30**, 514-537.
11. Bender, Lauretta. Gestalt functions in visual motor patterns in organic disease of the brain, including dementia paralytica, alcoholic psychoses, traumatic psychoses, and acute confusional states. *Arch. Neurol. Psychiat.*, 1935, **33**, 1-24.
12. Bender, Lauretta. A visual motor gestalt test and its clinical use. *Amer. Orthopsychiat. Ass., Res. Monogr.*, 1938, No. 3.
13. Bender, Lauretta. The Goodenough test in chronic encephalitis in children. *J. nerv. ment. Dis.*, 1940, **91**, 277-286.
14. Bender, Lauretta. Childhood schizophrenia. *Amer. J. Orthopsychiat.*, 1947, **17**, 40-56.
15. Bender, Lauretta. Psychological principles of

the visual motor gestalt test. *Trans. N.Y. Acad. Sci.*, 1949 **11**, 164-170.

16. **Bender, Lauretta, Curran, F. J., & Schilder, P.** Organization of memory traces in the Korsakoff syndrome. *Arch. Neurol. Psychiat.*, 1938, **39**, 482-487.

17. **Billingslea, F. Y.** The Bender-Gestalt: an objective scoring method and validating data. *J. clin. Psychol.*, 1948, **4**, 1-27.

18. **Cronbach, L. J.** *Essentials of psychological testing.* New York: Harper, 1949.

19. **Deri, Susan K.** *Introduction to the Szondi Test: theory and practice.* New York: Grune & Stratton, 1949.

20. **Fabian, A. A.** Vertical rotation in visual-motor performance—its relationship to reading reversals. *J. educ. Psychol.*, 1945, **36**, 129-154.

21. **Fehrer, E. V.** An investigation of the learning of visually perceived forms. *Amer. J. Psychol.*, 1935, **47**, 187-221.

22. **Foster, J. C.** Significant responses in certain memory tests. *J. appl. Psychol.*, 1920, **4**, 142-153.

23. **Gibson, J. J.** The reproduction of visually perceived form. *J. exp. Psychol.*, 1929, **12**, 1-39.

24. **Hanawalt, N. G.** Memory traces for figures in recall and recognition. *Arch. Psychol.*, 1937, **31**, No. 216, 1-89.

25. **Harrower, Molly R.** Neurotic depression in a child. In A. Burton and R. E. Harris (Eds.), *Case histories in clinical and abnormal psychology.* New York: Harper, 1947. Pp. 519-529.

26. **Harrower, Molly R.** Group techniques for the Rorschach test. In L. E. Abt & L. Bellak (Eds.), *Projective psychology.* New York: Knopf, 1950. Pp. 146-184.

27. **Hutt, M. L.** The use of projective methods of personality measurement in army medical installations. *J. clin. Psychol.*, 1945, **1**, 134-140.

28. **Hutt, M. L.** *A tentative guide for the administration and interpretation of the Bender-Gestalt Test.* U. S. Army Adjutant General's School, June, 1945 (restricted).

29. **Kadis, Asya L.** Finger-painting as a projective technique. In L. E. Abt & L. Bellak (Eds.), *Projective psychology.* New York: Knopf, 1950. Pp. 403-431.

30. **Machover, Karen.** *Personality projection in the drawing of the human figure.* Springfield, Ill.: Charles C Thomas, 1949.

31. **Mira, E.** Myokinetic psychodiagnosis: a new technique for exploring the conative trends of personality. *Proc. roy. Soc. Med.*, 1940, **33**, 173-194.

32. **Murphy, G.** *Personality: a biosocial interpretation.* New York: Harper, 1947.

33. **Napoli, P. J.** Finger-painting and personality diagnosis. *Genet. Psychol. Monogr.*, 1946, **34**, 192-230.

34. **Orenstein, L. L., & Schilder, P.** Psychological considerations of the insulin treatment in schizophrenia. *J. nerv. ment. Dis.*, 1938, **88**, 397-413.

35. **Pascal, G. R., & Suttell, Barbara J.** *The Bender-Gestalt Test.* New York. Grune & Stratton, 1951.

36. **Sapas, E.** Zeichnerische Reproduktionen einfacher Figuren durch Geisteskranke. *Schweiz. Arch. Neurol. Psychiat.*, 1918, **4**, 140-152.

37. **Sargent, Helen.** Projective methods: their origins, theory, and application in personality research. *Psychol. Bull.*, 1945, **42**, 257-293.

38. **Saudek, R.** *Experiments with handwriting.* New York: Morrow, 1928.

39. **Schilder, P.** Psychic disturbances after head injuries. *Amer. J. Psychiat.*, 1934, **91**, 155-188.

40. **Schilder, P.** Notes on the psychology of metrazol treatment of schizophrenia. *J. nerv. ment. Dis.*, 1939, **89**, 133-144.

41. **Wayne, D. M., Adams, M., & Rowe, L. A.** A study of military prisoners at a disciplinary barracks suspected of homosexual activities. *Milit. Surg.*, 1947, **101**, 499-504.

42. **Wechsler, D.** *The measurement of adult intelligence.* Baltimore: Williams and Wilkins, 1944.

43. **Wechsler, I.** *A textbook of clinical neurology.* Philadelphia: Saunders, 1947.

44. **Wertheimer, M.** Studies in the theory of gestalt psychology. *Psychol. Forsch.*, 1923, **4**, 301-350.

45. **White, R. W.** Interpretation of imaginative productions. In J. McV. Hunt (Ed.), *Personality and the behavior disorders.* Vol. 1. New York: Ronald, 1944. Pp. 214-251.

46. **Wolff, W.** Projective methods for personality analysis of expressive behavior in preschool children. *Charact. & Pers.*, 1942, **10**, 309-330.

47. **Wolff, W.** Graphometry: a new diagnostic method. *Psychol. Bull.*, 1942, **39**, 456. (Abstract)

48. **Wolff, W.** *The expression of personality.* New York: Harper, 1943.

49. **Wolff, W.** *The personality of the preschool child.* New York: Grune & Stratton, 1946.

50. **Wolff, W.** *Diagrams of the unconscious.* New York: Grune & Stratton, 1948.

51. **Woltmann, A. G.** The Bender Visual-Motor Gestalt Test. In L. E. Abt and L. Bellak (Eds.), *Projective psychology.* New York: Knopf, 1950. Pp. 322-356.

(Accepted for publication January 19, 1953)

Vol. 67, No. 7 Whole No. 357, 1953

Psychological Monographs: General and Applied

Psychological Changes During the First Year Following Prefrontal Lobotomy[1]

I. W. Scherer, J. F. Winne, D. D. Clancy, and R. W. Baker[2]
Veterans Administration Hospital, Northampton, Mass.

I. INTRODUCTION

Several authors surveying literature pertinent to the area of psychosurgery (9, 10, 11, 13, 18, 22) have generally concurred in three principal conclusions: first, that there is disagreement among experimenters regarding the effect which prefrontal lobotomy has on those functions which psychological tests are designed to measure; second, that certain important psychological variables have been neglected in the various studies; and third, that there is considerable discrepancy between interpretations of test data and clinical impression. The present writers, as a result of their own review of the experimental literature, wish to add a fourth criticism. It is apparent that several investigators have failed to consider basic methodological issues in evolving their experimental procedures. The fourth criticism is important, the writers feel, because it seems that much of the confusion in the field is due to inadequate experimental design. A brief discussion of five types of methodological weakness follows.

A major source of conflicting reports, and one common to many studies, is the matter of timing—i.e., *the length of the interval between pre- and posttestings.* Porteus and Peters (25) have pointed out the importance of having constant intervals between testing, since they have observed that the extent of functional deficit tends to decrease progressively following lobotomy. That is, the postoperative psychological condition is not static, but dynamic and changing. Consequently, an experiment which reports changes a few weeks following operation perhaps cannot be expected to be comparable with experiments which deal with longer or shorter time intervals.

---

[1] Abstract of this paper read at American Psychological Association, Chicago, September, 1951. From the Veterans Administration, Northampton, Massachusetts.

[2] The writers wish to acknowledge the help of the following people in the research study reported here:

Lionel M. Ives, M.D., Chief of Professional Services, who coordinated the lobotomy program in all its phases, selected the patients, and suggested certain areas for analysis; James C. Fox, Jr., M.D., who, as Consultant in Neurology and Psychiatry, made many thoughtful suggestions and criticisms; Roger Cheney, M.D., Consultant in Neurosurgery, who performed most of the operations, for his willingness to discuss what he was doing and why he was doing it; Robert Garrow, psychiatric aide, and Alphonse Sootkus, R.N., for their patience and cooperation; Howard Blane, John Byrne, Alfred Goldman, Herbert Lipton, Robert Misch, Melvin Rosenthal, and Howard Slepian, intern clinical psychologists, who assisted in the analysis of data and/or some of the testing; Leonard Kogan, Ph.D., Institute for Welfare Research, Community Service Society of New York, and J. Richard Wittenborn, Ph.D., Department of Psychology, Yale University, for statistical advice; Paul I. Yakovlev, M.D., Harvard University School of Medicine, for critical discussion of portions of the paper; Mrs. Ida Metz Hyland, Chief Social Worker, and other members of her department, particularly Miss Elizabeth L. Grover, for getting patients to return for their examinations when they were not in the hospital; Miss Eileen M. Kirby, departmental secretary, for maintaining schedules, recording data, carrying out numerous and tedious but necessary details, and typing this paper through all its many drafts; and Herbert S. Conrad, Ph.D., Editor of the *Psychological Monographs,* for his valuable assistance in preparing this manuscript for publication.

The second criticism holds that *the type of operation*, as an important variable, has not been adequately controlled. No matter how similar the various operative techniques may be, it is still not experimentally wise to interchange results when the patients have been subjected to alternative surgical procedures—e.g., lobotomy, lobectomy, topectomy, and gyrectomy. This is especially true when different areas of the cortex have been severed or removed. King's study of abstraction (20) in the Greystone Project illustrates the point, where differing results were found to be dependent upon the area ablated.

Many studies in the literature fail to supply the *control groups* which are necessary for establishing the base lines in terms of which changes due to lobotomy must be evaluated. Conclusions drawn from an experimental group, without first ascertaining that the changes observed are not due to some other variable than lobotomy, are at best tentative. Zubin (37), for example, found that Rorschach changes in control patients were often in the same direction as in the case of operated patients.

A fourth criticism has really two parts, both related to the problem of diagnostic classification. First, in some studies *the experimental groups are not homogeneous*—i.e., the lobotomized patients may include neurotics, psychotics, cases of brain damage, and cases of intractable pain. It cannot be assumed that lobotomy would have similar effects on such varied patients, since the preoperative conditions—i.e., the psychological organizations—are so disparate. For the same reason conclusions drawn from *homogeneous groups representing different diagnostic categories* should not be expected to be identical. For example, certain experiments have been interpreted to refute the findings of other experiments when in reality the "discrepant" studies have dealt with populations having different characteristics.

The fifth and last criticism refers to *the limited or atomistic viewpoint* which apparently determines the interpretation of psychological data in many studies. There has been little attempt made to integrate test findings with clinical data. Mettler (23) has pointed out the difficulties involved in such an approach and offers reasons for not having attempted it in the Greystone Project. It is the opinion of the authors, however, that unless all available data (from the case history, clinical observations, psychological tests, etc.) are integrated into a cohesive picture, only isolated information concerning the effect of lobotomy on behavior will be achieved. A methodological approach designed to answer this criticism will be described in the following section.

II. METHODOLOGY: SUCCESSIVE VERIFICATION OF HYPOTHESES

The experiment reported in this paper employed an extensive battery of psychometric and projective tests which was administered to the lobotomized patients about two weeks before operation, and two weeks, three months, and one year following. A group of control patients was tested at comparable intervals. The tests used and the selection of patients will be described in greater detail in succeeding sections. The remainder of this section offers a description of the methodology evolved by the authors—a threefold approach to the analysis of data.

Since statistical treatment alone cannot adequately characterize or portray changes which occur over the various time intervals, and since a purely qualitative consideration could not be sufficiently rigorous (for reasons given below), a method was developed which attempts to utilize both the quantitative and the qualitative aspects of the data. This approach appears to have several advantages over methods previously reported in the literature: (*a*) it permits consideration of individual differences which are sometimes masked by statistical techniques; (*b*) it minimizes the tendency sometimes found in qualitative analysis to prejudge data and to form *ad hoc* explanations; (*c*) it makes it possible to test qualitative deductions statistically; (*d*) it enables the experimenters to investigate more completely the pattern of behavioral changes during the postoperative period.

Although there are certain advantages to the type of qualitative approach used in this experiment, there are also some weaknesses. One aim of such an analysis was to develop a configurational picture of the patient, but in doing so the various components (test scores, history data, etc.) were neither equally weighted nor clearly defined. Moreover, the theoretical point of view espoused by the writers was a mixture of the thinking of Freud, Lewin, and some of the psychobiologists; it was not possible to be completely consistent in theoretical perspective when evaluating the results from successive time intervals, or even from one test to another within the same "time slice." Also, of course, the writers' biases—derived from extensive reading of the literature on brain damage in general, and on lobotomy in particular—have their effects on the treatment of data. Finally, it has proved extremely difficult, if not actually impossible, to keep in mind the configuration of hundreds of test scores and clinical items in terms of which the authors attempted to appraise each specific test change and its interpretative significance. Such an evaluation of test changes must take into account the alterations in the personality structure which occur as a result of (*a*) lapsed time, or those variables which are operative during the time lapse; (*b*) the altered interpersonal environment; and (*c*) changes in functions which are correlated with the specific function being examined. There is also, in the type of analysis used here, a tendency to subsume tests under broad categories in order to develop generalized rationales. The difficulty is that the rationale may be only more or less valid, thereby contributing to a false interpretation of the patient's psychological status. Finally, the last weakness pertains to the evaluation of scores on particular tests in a battery which is repeatedly administered. The tendency has been to interpret scores in the light of previous administrations of the battery, neglecting the interrelationships between measures from other tests. The authors feel that the significance of a score should be determined in terms of the matrix in which it occurs—i.e., in terms of the scores on the other tests of the battery at one particular administration. This aim, however, is most difficult, if not impossible, to achieve. One is usually compelled to treat merely the fluctuations which occur in each individual test from one administration to the next.

A more detailed description of our particular approach will now be attempted; it involved three stages of analysis.

The first step in the analysis of data was to formulate an integrated picture of the prelobotomy personality structure of each of a small sample of patients, and then to employ this unique configurational pattern as a frame of reference for evaluating postoperative changes. Prelobotomy test findings were integrated with all the other types of information concerning the patient, i.e., social history, psychiatric reports, ward observation notes, and behavior in the testing situation. (Excerpts from a forthcoming article dealing with this matter will be found in Appendix A.) These data were analyzed using two levels of inference. The first level involved primarily the study of ego intactness or reality testing; the second level dealt with the etiological variables within the personality constellation. Reality testing was defined in terms of performance on tests which are designed to reflect the efficiency of certain cognitive or ego functions such as attention, concentration, abstraction, memory, etc. Personality dynamics were formulated on the basis of the patient's performance and behavior in projective test situations. These interpretations were made in terms of various theoretical biases which were not as operationally determinable as those used with the first level data. Each test was then examined at the four time intervals in terms of the general configuration of ego functioning and the brief formulations of personality dynamics. Score changes were evaluated not only with respect to the single specific test, but also in relation to the pattern of changes in all tests of similar nature. From this analysis were derived tentative conclusions as to the nature of the changes which occur in the reorganization of personality following lobotomy. As mentioned, the foregoing detailed analysis was carried out on a small sample of the lobotomized group; a similar analysis, though not so exhaustive, was done with the other members of the experimental group.

The second step in the analysis of data, carried out by independent judges concurrently with the intra-individual analysis just described, was the longitudinal study of all patients on each single test. An example of this type of analysis will be found in Appendix B. The attempt here was to examine the change in the lobotomy group over the total period of time, i.e., over the course of one year. This involved the determination of those changes which occurred in each population (experimental and control) at the various time intervals. This test-by-test longitudinal analysis afforded impressions as to change or lack of change in variables reflecting adjustment in such areas as motor, emotional, social, etc. Those impressions which were in accord with the tentative conclusions derived from the intra-individual analysis were retained as working hypotheses for the third phase of the analysis.

The third phase of the analysis consisted of statistical tests of the four hypotheses derived from the individual case studies and the longitudinal analyses. Tests were made of the statistical significance of the net change in means for the experimental (lobotomized) group for each of six time periods: (*a*) preoperation to two weeks post; (*b*) pre to three months post; (*c*) pre to one year post; (*d*) two weeks post to three months post; (*e*) two weeks post to one year post; and (*f*) three months post to one year post. Since not all subjects could be tested on all variables at each occasion, either because of lack of cooperation or unavailability due to illness, trial visit, etc., the sample of patients varies from a maximum of 22 to a minimum of 13.[3] Also, because the same subjects could not be fully tested on each occasion, the data utilized in the statistical analyses are based on a sample which varies in maximum size from 22 to 12. Since only nine subjects in each group took all the tests on each occasion, the statistical analyses are based on data from those individuals who took a given test on at least two occasions.

The existence of incomplete data, the small number of subjects, the small number of administrations of the tests, and the large number of variables (which are not all independent) precluded the use of factor analytic methods. Further, the lack of normality of the distributions of scores and the lack of homogeneity of variance suggested that analysis of variance or covariance was not feasible. It was decided, therefore, to test the statistical significance of the *net change* in the experimental group, in spite of the untenable basic assumption that initial and final scores have a one-to-one relationship.

The formula employed for testing net change, using raw scores and small sample techniques, is as follows:

$$t=\frac{\dfrac{\Sigma(Y-X)}{N}-\dfrac{\Sigma(Y'-X')}{N'}}{\sqrt{\dfrac{N\Sigma(Y-X)^2-[\Sigma(Y-X)]^2}{N^2(N-1)}+\dfrac{N'\Sigma(Y'-X')^2-[\Sigma(Y'-X')]^2}{N'^2(N'-1)}}}$$

where X and X' are initial raw scores in the experimental and control groups, Y and Y' are final raw scores in the experi-

---

[3] Tables giving the number of cases, mean, and standard deviation for each variable for the total sample have been deposited with A.D.I. For tables order Document 3908 from American Documentation Institute, c/o Library of Congress, Washington 25, D.C., remitting $2.25 for microfilm (images 1 inch high on 35 mm. motion picture film) or $5.00 for photostats readable without optical aid.

TABLE 1
DESIGN OF STATISTICAL ANALYSIS

Experimental				Control			
1st test	2nd test		Total	1st test	2nd test		Total
	Below	Above			Below	Above	
Above Below	A C	B D	$A+B$ $C+D$	Above Below	A' C'	B' D'	$A'+B'$ $C'+D'$
Total	$A+C$	$B+D$	N	Total	$A'+C'$	$B'+D'$	N'

$A, B \cdots D'$ are cell frequencies, $Z = \frac{N'(|D-A|-1)-N(|D'-A'|-1)}{\sqrt{N'^2(A+D)+N^2(A'+D')}}$ · Z is evaluated from entries in the normal probability table.

mental and control groups, $df = N + N' - 2$.

Although the distributions were skewed, it was not felt that transformations of raw scores were warranted in this experiment. By using raw scores the chances of obtaining statistical significance were actually reduced, since the standard error of the difference between means was inflated relative to the displacement of the means. It was felt, therefore, that if significant results were obtained with skewed data, nothing would be gained through transformation. The few variables for which a true difference in means seemed to be masked by inflation of the standard error were dichotomized by splitting the obtained distributions at the median of the combined distribution of preoperation test scores and tested as described in the next paragraph.

Dichotomized data were tested for the significance of the net shift in proportions by a formula adapted by Kogan[4] from McNemar (21) which corrects for continuity. Fourfold tables were developed for each group (Table 1).

This paper deals primarily with the final stage of the threefold analysis and reports the results of the statistical tests of the four hypotheses which were derived from the extensive case material and the longitudinal study of tests. The authors realize that their method of successive verification of hypotheses does not constitute cross validation in the usual sense, since the same total population was used throughout. There were differences between samples used in the three stages of analysis, however, which make for reduction of contamination. Intensive case reports were promulgated for only two or three lobotomized patients, although all experimental records were briefly surveyed. Longitudinal analyses of tests are based on the total sample of patients tested on any single occasion. Finally, the statistical analysis is based on those individuals who were tested on at least two occasions. The fact that each judge in the first two phases of the analysis worked independently, without knowledge of his co-workers' results, also serves to reduce *ad hoc* thinking.

III. EXPERIMENTAL PROCEDURE

A. SUBJECTS

Forty-four subjects were used in this experiment, all of whom were psychotic male patients in the Veterans Administration Hospital, Northampton, Massachusetts. Twenty-two of these patients, the experimental group, underwent bilateral prefrontal lobotomy. A control group of 22 patients was fairly well equated with the operated group for age, sex, education, length of hospitalization, cooperation level, diagnosis, and certain behavior characteristics (e.g., assaultiveness, talkativeness, etc.).

---

[4] Personal communication. January 4, 1950.

The experimental group averaged 33.9 (*S.E.* 1.3) years of age, had received 10.8 (*S.E.* 0.5) years' education, and had been hospitalized for 4.09 (*S.E.* 0.5) years. The control group averaged 30.4 (*S.E.* 1.0) years in age, had received 10.05 (*S.E.* 0.7) years' education, and had been hospitalized 3.3 (*S.E.* 0.4) years. Fourteen experimental and eleven control patients manifested an unusually high level of cooperation. The experimental group was made up of five paranoid schizophrenics, eight catatonics, four hebephrenics, and five schizophrenics, mixed type. The control group consisted of four paranoid schizophrenics, eight catatonics, two hebephrenics, and eight schizophrenics, mixed type. All 44 patients had been approved for lobotomy by a board of psychiatrists following an unsuccessful course of electroconvulsive shock treatment, and were receiving the same types of adjunctive therapy.

The authors wish to point out that the control conditions were not as complete as they would like, since the control patients did not undergo a "mock" operation.

B. Procedure

Twenty-nine tests comprised the psychological battery which was employed in this experiment, from which 145 variables were derived. Most of the variables were selected before testing was begun; all of them were selected before any analysis was commenced. The battery was administered to the experimental group about two weeks before, and two weeks, three months, and one year following bilateral prefrontal lobotomy. The control group was tested at comparable time intervals.

In order to define more clearly the type of operative procedure that was used, the following excerpt is quoted from the neurosurgeon's report of a typical standard, open Lyerle-Poppen prefrontal lobotomy.

> Under local pentothal anaesthesia, the usual skin incisions were placed in plane of pupil parallel to the sagittal sinus 3½ cm. from the midline. Trephine openings were made 2 cm. anterior to the coronal suture and the bone buttons removed. Transverse cortical incision was made and the sphenoid ridge identified by the use of the brain cannula. The plane of section was then made through the frontal lobes passing 1 cm. anterior to the tip of the frontal horn bilaterally. There was no difficulty with hemostasis. The ventricles were not encountered. Silver clip markers were placed and the wound irrigated.

The tests were administered in group fashion whenever practicable. Those tests which have published manuals were scored according to the prescribed method; the evaluation of qualitative data was achieved by the group judgment of three clinical psychologists.

Analyses, as described in the preceding section on methodology, were made for six time periods: (*a*) preoperation to two weeks post; (*b*) pre to three months post; (*c*) pre to one year post; (*d*) two weeks post to three months post; (*e*) two weeks post to one year post; and (*f*) three months post to one year post.

C. Test Materials

A complete list of tests and variables follows.

1. *Digit Symbol.*

Subtest 10 from the Wechsler-Bellevue Scale of Adult Intelligence (34). Administered group fashion with Wechsler's directions, time limit, and weighted scores.

(1) Errors, number of
(2) Weighted score

2. *Digit Span.*

Subtest 4 from the Wechsler-Bellevue Scale

(34). Administered individually with standard directions and scoring.

(3) Digits forward, number of
(4) Digits reversed, number of
(5) Weighted score

3. *Successive Subtraction.*

Oral serial subtraction of 7 from 100 (35), given individually. Adding instead of subtracting, or subtracting 10 instead of 7 are interpreted as rigidity.

(6) Time, in seconds
(7) Errors, number of
(8) Rigidity, presence-absence

4. *Hard Pairs.*

Subtest 3 from the Hunt-Minnesota Test for Organic Brain Damage (16); a list of 10 word pairs. Given individually with a single reading of the paired words and a single test for the number of correct associations to the stimulus word in each pair.

(9) Correct associations, number of

5. *Visual Memory.*

The Benton Visual Memory Test (4); geometric designs to be reproduced from memory, individually administered. In addition to the number of correct reproductions, certain formal aspects of the drawings were scored.

(10) Correct designs, number of
(11) Size, rated on 5-point scale
(12) Retracing of lines, presence-absence
(13) Difficulty with reproduction of angles, presence-absence
(14) Heaviness of lines, presence-absence
(15) Lightness of lines, presence-absence
(16) Waviness of lines, presence-absence

6. *Vocabulary.*

Subtest 5A from Wechsler-Bellevue Scale (34). Given group fashion; patients were asked to write meanings of words printed on test sheet. Standard scoring procedure used. The derived score was used as a base line against which to compare performance on the Memory Paragraph and on the Shipley-Hartford Scale.

(17) Weighted score

7. *Memory Paragraph.*

Test 6 from Babcock-Levy battery (2), administered in group fashion by having patients write whatever they recall of an orally presented paragraph. There were two recalls, the sequence being: (*a*) reading of stimulus material to group; (*b*) immediate written recollection; (*c*) re-reading of the paragraph; (*d*) 10 minutes of interpolated testing using other nonrelated materials; (*e*) another written recall. A memory efficiency score was obtained by forming a ratio of the average number of correct memories on the two recalls to the vocabulary weighted score.

(18) Immediate correct memories, number of
(19) Distortions of content in immediate recall, number of
(20) Distortions of order in immediate recall, presence-absence
(21) Correct memories in delayed recall, number of
(22) Distortions of content in delayed recall, number of
(23) Distortions of order in delayed recall, presence-absence
(24) Average number of correct memories on the two recalls
(25) Average number of distortions of content on the two recalls
(26) The memory efficiency score, described above
(27) Number of correct memories in delayed recall minus the number in immediate recall
(28) Number of distortions in delayed recall minus the number in immediate recall

8. *Finger Dexterity.*

An adaptation of Johnson O'Connor's Worksample #16 (24), using the tweezer dexterity plate instead of the one prescribed. Administered individually, patients are required to place 50 pins, using the fingers of the dominant hand, in 5 rows of 10 pins each.

(29) Time, in seconds

9. *Dynamometer.*

A standard hand dynamometer employed in a situation designed to yield a measure of level of aspiration. Given individually, the patient is asked to squeeze the apparatus as hard as possible. After being informed of his score, he is asked to state how hard he thinks he can squeeze it on a second trial.

(30) Direction of goal (up, down, same)

10. *Tweezer Dexterity.*

An adaptation of Johnson O'Connor's Worksample #17 (24). The trials are run in an individual situation. The patient is required, using a tweezer in his dominant hand, to place 30 pins in 3 rows of holes, 10 in each row. The patient is asked to predict his performance on the second trial from a knowledge of his performance on the first trial.

(31) Total time, in seconds, for the two trials
(32) Direction of goal

11. *Steadiness Test.*

An adaptation of Halstead's Manual Steadiness Test (14), individually administered, where the patient's task was to run through two trials of placing a stylus in a series of progressively smaller holes, trying not to touch the sides. The patient is asked to predict his performance on the second trial from a knowledge of the first.

(33) Total number of contacts on the two trials
(34) Direction of goal

12. *Downey Inhibition Test.*

Subtests VI-2 and VII from the Downey Group Will-Temperament Test (7), given in a group situation. The former task requires the subject to write "United States of America" as rapidly as possible; in the latter the patient is asked to trace a scroll of dotted lines as slowly as possible.

(35) Time for USA, in seconds

(36) Total time for tracing scroll, in seconds, for 3 trials (interpreted as measure of inhibition)

(37) Difference, in seconds, trial 3 minus trial 1

13. *Halstead Tactual.*

Halstead's modification of the Seguin-Goddard Formboard (14). The patient, blindfolded, was required to place ten variously shaped blocks in their appropriate holes on a board erected 70 degrees from the horizontal. There were three trials—right hand, left hand, both hands—after which the patient was asked to reproduce the formboard from memory with paper and pencil. A measure of learning was obtained by determining the time difference between the first and succeeding trials; certain formal aspects of the drawings were also scored. Individually administered.

(38) Right hand, time in seconds

(39) Left hand, time in seconds

(40) Both hands, time in seconds

(41) Difference between right and left hands, in seconds

(42) Difference between right and both hands, in seconds

(43) Shapes correctly drawn, number of

(44) Shapes correctly localized, number of

(45) Perseveration of shape, presence-absence

(46) Heavy lines, presence-absence

(47) Light lines, presence-absence

14. *Object Sorting.*

The test materials, for the most part, corresponded to those described by Goldstein and Scheerer (12) and Reichard and Rapaport (27). Administered group fashion, the objects were placed in the center of a table in full view of the seated patients. Seven stimulus objects were successively presented by the examiner, and the patients were asked to write the name of each and the names of any objects on the table which could belong with each. The patients were also requested to write the reasons for their various sortings. Scoring was based on qualitative evaluation of the active sortings discussed in Rapaport, Gill, and Schafer (26), Goldstein and Scheerer, and from the clinical experience of the senior author. Each sorting was scored once, on a scale from 0 to 3, for correctness or adequacy of the choice of items with reference to the stimulus object, and a second time for the adequacy of the concept or abstraction given.

(48) Total score

(49) Confabulation, presence-absence

(50) Symbolism, presence-absence

15. *Similarities.*

This subtest from the Wechsler-Bellevue Scale (34) was given group style by having the patients write their responses. Standard directions and scoring were used.

(51) Weighted score

16. *Series Completion.*

A speed test from the Army Wechsler (1), given group fashion, where the patient's task is to complete a series of geometric figures by abstracting the relationship which is represented.

(52) Attempted solutions, number of

(53) Correct responses, number of

17. *Shipley-Hartford.*

A subtest from the Shipley-Hartford Retreat Scale (32), similar to Series Completion, but dealing with verbal material. Administered group style. The conceptual quotient, one of Shipley's measures, was determined by finding the ratio of the score on this test of abstraction to the Wechsler vocabulary score.

(54) Attempted solutions, number of

(55) Raw score (twice the number of correct responses)

(56) Conceptual quotient

18. *Categorization.*

This test, originally developed by H. M. Capps (5), was given group fashion. There are 25 test items, each consisting of a stimulus word followed by seven other words. The patient's task is to underline the three words in each item which are related to the stimulus word.

(57) Raw score (the number of correct items)

19. *Trail Making.*

From the Army Individual Test, Form I (38). Although two forms of the subtest, A and B, were given to the patients, only the results of Form B were evaluated. Group administered, the subject's task was to connect successively by drawn line twenty-five randomly spaced circles designated by numbers or letters. Directions and scoring are consistent with standard procedure.

(58 Time, in seconds

(59) Errors, number of

20. *Draw A Woman.*

Given group fashion; the patients were asked simply to draw a woman. The rationale for the selection of particular scoring categories—largely defined in terms of formal characteristics of the drawings—was based principally on unpublished research by Lydia Levy.[5]

(60) Size (on 5-point scale)

(61) Breast size (5-point scale)

(61a) Shift in breast size from one testing to another

(62) Confusion of sex, presence-absence

21. *Draw-A-Man.*

What was said about the Draw-A-Woman Test also applies here, except that the patient was asked to draw a man.

(63) Size (5-point scale)

(64) Nose size (5-point scale)

(65) Accuracy or degree of resemblance to a human figure (5-point scale)

(66) Confusion of sex, presence-absence

---

[5] Personal communication, 1946.

(67) Head or full-figure representation
(68) Discontinuous lines ("chicken scratching"), presence-absence
(69) Shading (the solid coloring, by pencil, of sizable portions of the figures), presence-absence
(70) Retracing of lines, presence-absence
(71) Heavy lines, presence-absence
(72) Light lines, presence-absence

22. *Bender-Gestalt.*

This test was individually administered by the standard method (3). Formal aspects of the drawings were analyzed and scored; the signs of regression used are those described by Bender (3), Hutt (17), and others.

(73) Size (5-point scale)
(74) Discontinuous lines, presence-absence
(75) Elaboration (the addition of extraneous lines), presence-absence
(76) Retracing of lines, presence-absence
(77) Difficulty with angles, presence-absence
(78) Heavy lines, presence-absence
(79) Light lines, presence-absence
(80) Regression signs, presence-absence

23. *Word Association.*

The word list was that developed by Rapaport, Gill, and Schafer (26). Given group style, the patients were instructed to read the stimulus words and to write beside each one the first word that came to mind. There were two pages, each having thirty words. Upon completion of the task the patients were given another copy of the test and asked to write the same words which they had written the first time.

(81) Percentage of correct recall
(82) Number of omissions (stimulus words not responded to) for first presentation of the 60 words
(83) Phrase responses, number of (first presentation)
(84) Close responses, number of (first presentation)
(85) Distant responses, number of (first presentation)
(86) Clang responses, number of (first presentation)
(87) Personal references, number of (first presentation)
(88) Time, in seconds, for first thirty words

24. *Picture Frustration.*

Group administration by standard procedure of this technique developed by Rosenzweig (30). Only the extrapunitive responses were scored.

(89) Extrapunitive ego-defensive responses, number of

25. *Szondi.*

A single administration of this test was given in a group situation, as opposed to more orthodox methods of administration (6). The pictures were placed on a table in full view of the patients, and each card in a series was identified by a number from 1 to 8. Each patient was given a mimeographed form and instructed to designate those two pictures he liked by placing an "L" next to the appropriate numbers on the form. Two disliked pictures in each of the eight series were also to be chosen and so indicated by writing a "D" by the corresponding numbers. Positive (liked), negative (disliked), and total choices for each vector were scored.

(90-92) *h* choices: liked, disliked, total
(93-95) *s* choices: liked, disliked, total
(96-98) *e* choices: liked, disliked, total
(99-101) *hy* choices: liked, disliked, total
(102-104) *k* choices: liked, disliked, total
(105-107) *p* choices: liked, disliked, total
(108-110) *d* choices: liked, disliked, total
(111-113) *m* choices: liked, disliked, total

26. *Attitude toward God.*

Adapted from Thurstone and Chave's Scale No. 22 (33), administered group fashion. The test consists of 20 statements pertaining to the concept of God; the patients were asked to make a choice of agreement or disagreement with regard to each one. Only those items on which the patient expressed agreement were scored, and these in turn were broken down as to whether they were statements of affirmation or of negation.

(114) Number of affirmative statements agreed with
(115) Number of negative statements agreed with

27. *Aversion Test.*

This is an unpublished scale of 26 items which was constructed by the senior author in collaboration with Dr. James Gibson, its purpose being to reveal attitudes towards certain asocial activities. Scores for each item range from 0 ("would not bother me at all") to 4 ("impossible to consider thinking or doing" the particular activity). Group administered.

(116) Attitude toward killing someone (without danger of punishment)
(117) Attitude toward keeping money found in a purse
(118) Attitude toward staying at home for two months
(119) Attitude toward incestuous thoughts
(120) Attitude toward deviant sexuality (pederasty)
(121) Attitude toward staying at home for one week
(122) Attitude toward masturbation in daytime
(123) Attitude toward eating cooked human flesh (under pressure of hunger with no reprisal)
(124) Attitude toward urination on public street during daytime
(125) Total aversion score (on the 26 items)

28. *Individual Rorschach Examination.*

The Rorschach test was individually administered using standard directions (29). No inquiry was made, and relatively little encouragement was given, since in this phase of Rorschach examination the authors were primarily concerned with reaction time and rejections.

(126) Rejections, number of
(127-136) Reaction time on each of the 10 cards
(127) Average reaction time, excluding rejected cards

29. *Multiple-Choice Rorschach.*

Given individually immediately following the individual Rorschach, using the same plates. Subjects were asked to pick, from ten choices for each card, the one which best described the plate in question. The choices for card V were taken from Form A of the Harrower-Erickson technique (15); for the other nine cards Form C was used. The poor anatomy responses (F_o) were scored according to the Harrower-Erickson-Steiner manual. Those responses with movement and color determinants were scored on the basis of the recommendations in the manual, the senior author's clinical experience, and the majority agreement of three independent scorings made by staff clinical psychologists.

(138) Rejections, number of
(139) $FC + CF + C/F$ responses, sum of
(140) $C_n + C_{des} + C_{sym}$ responses, sum of
(141) C responses (not including those in 140), number of
(142) Poor anatomy responses (F_o), number of
(143) $H + Hd$ responses, sum of
(144) Sex responses, number of
(145) Shading responses, presence-absence

IV. RESULTS

It will be remembered that in the section on methodology, two types of data analysis were described by means of which hypotheses were derived that could, in turn, be subjected to statistical test. Four major hypotheses were so developed, and will be individually discussed in the following paragraphs.

Hypothesis I

The first hypothesis states that there is a predictable pattern of performance on tests related to organic factors subsequent to the lobotomy operation: immediately following the surgical procedure there will be a decrease in mental efficiency; as recovery from the operation takes place, test performance will return to the preoperative base line or rise slightly above it, reaching a peak at three months postoperation; finally there will be a general decline such that the status one year after operation will be little better than that prior to surgery.

Thirty-nine relatively objective variables believed to measure mental efficiency were selected by the authors and examined as to the statistical significance of the net change in the lobotomized group. The direction of change was predicted by the senior author in advance of this analysis, thereby justifying a one-tailed test of significance. Table 2 gives the predicted change and the observed direction of change for each of the 39 variables at the six time periods. Of the possible 234 changes, we note that there are 68 observed shifts at the .05 level of significance or better, with 54 shifts in the expected direction (either "H" or "L"), while by chance we expect a total of only 23.4 (at the .05 level of significance) to be in the expected direction. The significant results obtained from the table as a whole ($p < .01$) indicate that the qualitative approach used in setting up hypotheses has considerable merit. At each time period, we would expect by chance alone to have 3.9 changes at the .05 level (10% of 39) in either direction. Thus, our results seem to be significant for the pre to two weeks post, two weeks post to three months post, and the two weeks post to one year post comparisons ($p < .01$). There is some doubt as to the validity of the pre- to one year postoperation comparison, since the probability associated with our results is greater than .05 and less than .10. Significant results were not obtained for the pre to three months post and three

TABLE 2

DIRECTION OF NET CHANGE IN VARIABLES RELATED TO HYPOTHESIS I: ORGANICITY, MENTAL EFFICIENCY

(H indicates higher mental efficiency [decreased organicity]; L indicates lower mental efficiency [increased organicity]; — indicates no change in either direction)

Variable	Time Period					
	Pre–2 wk.	Pre–3 mo.	Pre–1 yr.	2 wk.–3 mo.	2 wk.–1 yr.	3 mo.–1 yr.
Expected Change	L	H	H	H	H	L
Organic Signs						
Word Assoc Phrases (83)	L	—	—	—	H	—
Word Assoc Close (84)	L	—	—	H	—	—
Word Assoc Pers Ref (87)	—	—	H	—	H	—
D-A-M Shading (69)	L	—	—	—	H	—
D-A-M Accuracy (65)	—	H	—	H	—	L
B-G Heavy Lines (78)	—	—	—	—	H	—
B-G Elaboration (75)	—	—	—	H	H	—
B-G Retracing (76)	—	—	—	—	H	—
Vis Memory Retracing (12)	—	—	—	H	—	—
Vis Memory Heavy Lines (14)	—	—	—	H	—	L
Abstraction						
Categorization (57)	L	—	—	H	H	—
Similarities (51)	—	—	—	—	—	L
Shipley Rights (55)	—	—	H	—	—	—
Shipley C.Q. (56)	—	—	—	L	—	—
Object Sorting (48)	L	—	—	—	—	—
Motor						
Tweezer Dexterity (31)	L	—	—	H	—	L
Finger Dexterity (29)	—	H	—	H	—	—
Steadiness Test (33)	L	—	—	—	—	—
Downey Total (36)	L	L	—	H	H	H
Downey T.3–T.1 (37)	—	—	—	—	—	L
Halstead Right Hand (38)	L	—	L	—	—	—
Motor Memory						
Halstead Both Hands (40)	L	H	—	H	—	—
Halstead Right-Left (41)	—	—	H	—	—	—
Halstead Right-Both (42)	—	H	—	—	—	—
Memory						
Visual Memory (10)	—	H	—	H	—	—
Hard Pairs (9)	—	—	H	—	—	—
Mem Par Immed Rights (18)	—	—	H	—	—	—
Mem Par Delay Rights (21)	—	—	H	—	—	—
Mem Par Average Rights (24)	—	—	H	—	—	—
Mem Par Delay Distort (22)	—	—	—	—	H	—
Mem Par Average Distort (25)	—	—	—	—	H	—
Digit Span Forward (3)	—	L	—	—	—	—
Digit Span Reverse (4)	—	L	—	L	—	—
Digit Span Wghtd Score (5)	L	L	—	—	—	—
Cognitive						
Digit Symbol Score (2)	—	H	—	H	—	—
Digit Symbol Errors (1)	—	—	—	—	—	H
Vocabulary (17)	—	—	—	—	H	—
Trail Making Errors (59)	—	—	L	—	—	L
Serial Seven Errors (7)	—	L	H	L	L	H
Changes in Expected Direction	11	6	8	12	11	6
Changes in Opposite Direction	0	5	2	3	1	3

months post to one year post comparisons.

At the two-week period, operated patients show less efficiency in eight of the tests found in Table 2 under the subdivisions of motor ability, cognitive efficiency, abstraction, and memory. There are also significant changes in three of the tests purported to reflect the effects of brain damage. At three months, however, the situation is almost reversed. With the exception of memory for digits forward and reversed, errors on successive subtraction, and inability to inhibit motoric action (Downey), these patients show either a return to, or an increase over, the preoperative efficiency level. The upward trend is most evident in the comparisons between the two-week–three-month period, and the two-week–one-year period. The authors believe that the exceptions to the general upward trend suggest the presence of permanent damage to the brain. At one year postoperation, many of the measures are still above the preoperative base line, but there appears to be a trend from three months to one year in the direction of decreased mental efficiency.

It may be concluded then, in support of the first hypothesis, that performance on the following kinds of tests seems to be influenced by prefrontal lobotomy: abstraction, pursuit of goals, memory, intellectual energy, accuracy of the body image, attention, concentration, and neuromuscular coordination. It is possible that these fluctuations reflect changes in the anatomical structure of the brain. That is, the initial decline may be associated with the lesion resulting from the surgical assault and then, as the lesion heals, the patient also recovers in terms of mental efficiency. There is evidence that pathological tissue changes occur even after the original lesion is "healed," which may account for the later decline. If, then, the changes in psychological function are in fact associated with changes in brain structure, they should be reflected in those tests designed to measure mental efficiency, as well as those tests commonly assumed to indicate the presence of organicity (organically caused malfunctioning).

B. Hypothesis II

The second hypothesis is that following lobotomy there is an increase in the strength of ego boundaries.[6] In order to test this hypothesis the senior author selected 18 test variables—some related to personality and some to mental efficiency—which are assumed to reflect degree of contact with the environment, and predicted the direction of change at the several time periods. The results of the statistical examination are presented in Table 3, where the observed direction of change for the 18 variables is indicated for the six time periods. Of 108 possible changes, there were 41 obtained at the .10 level of significance or better, with 34 of them in the expected direction. This over-all result is associated with a probability less than .01. Tests of significance for individual time periods suggest that our distributions differ from chance (in the expected direction) for all comparisons except the three months post to one year. Some tendency toward increased ego boundaries at the two-week posttesting is noted, as indicated by a probability value between .10 and .20.

That postoperative meaningful memory and learning ability show improvement over the pretest levels, as manifested in six test variables, is clearly seen in Table 3. An observed shift toward increased accuracy also is found in the Draw-A-Man test, reaching a peak at three months post. The shift contrary to predicted direction in this

---

[6] Loss of ego boundary is defined as an experience "in which the discrimination between the consciousness of self and the consciousness of the object were (*sic*) entirely suspended, the ego being no longer separated from the nonego; the subject no longer distinct from the object; the self and the world were fused in an inseparable total complex." Quoted in *Psychiatric Dictionary* (Hinsie, L. E. and Shatzky, J. New York: Oxford University Press, 1940, p. 189) from Storch, A. *The Primitive Archaic Forms in Schizophrenia.* (Tr. by Willard, C.) New York and Washington: Nervous and Mental Disease Publishing Company, 1924.

TABLE 3

DIRECTION OF NET CHANGE IN VARIABLES RELATED TO HYPOTHESIS II: INCREASED EGO BOUNDARIES

(H indicates increased, L indicates decreased, and — indicates no change in ego boundaries)

Variable	Time Period					
	Pre-2 wk.	Pre–3 mo.	Pre–1 yr.	2 wk.–3 mo.	2 wk.–1 yr.	3 mo.–1 yr.
Expected Change	H	H	H	H	H	H
Memory						
Digit Span Forward (3)	L	L	—	—	—	—
Mem Par Immed Rights (18)	H	—	H	—	—	—
Mem Par Delay Rights (21)	—	—	H	—	—	—
Mem Par Average Rights (24)	—	—	H	—	—	—
Cognitive (Learning)						
Digit Symbol Score (2)	—	H	—	H	H	—
Halstead Right-Left (41)	—	—	H	—	—	—
Halstead Right-Both (42)	—	H	—	—	H	L
Vocabulary (17)	—	H	H	H	H	—
Body Image						
D-A-M Accuracy (65)	—	H	—	H	—	L
Contact Variables						
Rorschach Rejections (126)	H	—	—	—	—	H
Mult Choice Ror Reject (138)	H	—	—	—	—	—
Mult Choice Ror C Resp (141)	H	H	H	—	—	—
Mult Choice Ror C_n Resp (140)	—	—	—	H	—	—
Szondi p+ (105)	—	—	—	H	—	L
Szondi p− (106)	H	H	H	—	—	L
Word Assoc Pers Ref (87)	—	H	H	H	H	—
Dynamometer Level of Asp (30)	H	—	H	—	—	—
Tweezer Dext Lev of Asp (32)	—	—	—	H	—	L
Changes in Expected Direction	6	7	9	7	4	1
Changes in Opposite Direction	1	1	0	0	0	5

variable at the three-month–one-year period will be dealt with later. The changes noted in digit span forward, also opposite to the expected direction, may not be inconsistent with the hypothesis. That is, it is possible that a perseverative factor, a concomitant of weakened ego boundaries, was operative at the presurgical level. The decreased score postoperation may reflect a diminution of the perseverative tendency which (if it is, as suggested, negatively correlated with the strength of ego boundaries) may be associated with the ability to differentiate between self and non-self.

Observed changes also compare quite favorably with predicted shifts in the contact variables listed in the table. On the Rorschach, there is a tendency to reject fewer cards, to give more human content, and to offer fewer crude color responses. Interpretively these changes would indicate that the patient has more energy to invest in environmental objects, is better able to identify with human figures in the environment, is less controlled by impulsive emotional forces, and displays a somewhat greater inclination to give commonly acceptable percepts. The incidence of personal references on the Word Association Test, responses indicative of autistic thinking, decreases. The scores of two Szondi signs *p*+ and *p*− shift according to expectancy, which would seem to indicate an increased orientation toward the environment or a willingness to make contact with the outside world, with less withdrawal and projection. All of the results discussed in this paragraph may be considered as signs of increased environmental contact.

A tendency toward a higher vocabulary score suggests increased functional intelligence, and there is also a tendency to set higher goals in the tests of level of aspiration (dynamometer, finger dexterity). Related to these variables is the willingness of the patient to expend mental energy on simple tests. This was manifested by an increased number of attempted solutions on the series-completion test in the two-week–three-month period. (Series completion is not included in Table 3 because it was not one of the tests on which changes were predicted to occur.) However, on a corresponding measure for the Shipley-Hartford, where a similar but more difficult task is involved, the patients appear to be less willing to expend this mental energy. That is, in the two-week–three-month period, there is a decrease in attempts on the Shipley-Hartford at the .01 level.

The authors believe that the foregoing results tend to confirm the hypothesis that ego boundaries are strengthened following prefrontal lobotomy. It will be noted in Table 3, however, that the observed shifts at the three-month–one-year period tend to be contrary to the predicted direction. While these results do not controvert the hypothesis they do necessitate modification. It appears that the strengthening of ego boundaries after operation is not a progressive process culminating in a final state of sustained improvement; rather, the improvement reaches a peak and then a trend toward decline, similar to that noted in the discussion of the first hypothesis, sets in.

The functional similarity between variables measuring organically determined malfunction and the variables measuring ego functions makes it difficult, in a clinical analysis, to separate them. Thus, we predicted more efficiency at the three months posttesting in certain tests related to the first hypothesis, but were unable to support this hypothesis. However, we do note that on tests of ego functioning related to the second hypothesis, improvement does take place. On the other hand, the predicted strengthening of ego boundaries does not hold up in the period from three months to one year postoperation, perhaps because of decreased performance on variables related to the organic hypothesis. While the individual clinician may be able to predict over-all efficiency at various time limits, statistical analysis appears to be a better way to identify the mechanisms which seem to be responsible for the gross fluctuations.

In order to explain dynamically the alteration in ego structure, it is the opinion of the authors that lobotomy represents an attack upon the body image or schema, or ego, which at first seems to accentuate the schizophrenic regressive process. In the course of this regression, a hypercathexis of ego boundaries occurs—a defensive constriction of the self—that serves to make sharper distinction between that which is ego and that which is nonego. This opinion is consistent with the thinking of Frank (8) and Reitman (28).

C. Hypothesis III

The third hypothesis to be examined is that lobotomy results in an increased sexual awareness and an increased differentiation, in the patient's mind, between masculinity and feminity. To test this hypothesis the senior author selected and predicted the direction of change in eight variables which, from his experience and from clinical opinion appearing in the literature, would seem to be related to awareness of sexuality. The results of the statistical analysis are found in Table 4, where the observed direction of shift in the eight selected variables is indicated for the six time periods. Of the 48 possible shifts, we observe 18 at the .10 level or better, with 16 in the expected direction. This over-all result is presumed to be significant at the .01 level. However, since only the pre to two-week and the pre to one-year results are significant by themselves (with a probability less than .10) the support for the hypothesis can be considered to be only tentative.

In support of the hypothesis the results show decreased sexual confusion on the Draw-A-Man test, less difficulty with angles on the Bender, and a decreased reaction time to Rorschach card VI. There is an increase in the size of the breast on the female figure and an improved conception of the body image (more accurate drawing of the man). Greater concern over sexuality is demonstrated in increased aversion toward masturbation, and there is a tendency toward a higher incidence of sex responses on the multiple-choice Ror-

TABLE 4

Direction of Net Change in Variables Related to Hypothesis III: Sexual Awareness and Differentiation

(H indicates more, L indicates less, — indicates no change in sexual awareness)

Variable	Time Period					
	Pre– 2 wk.	Pre– 3 mo.	Pre– 1 yr.	2 wk.– 3 mo.	2 wk.– 1 yr.	3 mo.– 1 yr.
Expected Change	H	H	H	H	H	H
Aversion to Masturbation (122)	H	—	—	—	—	L
B-G Angle Difficulty (77)	H	—	H	—	—	H
D-A-M Sexual Confusion (66)	—	H	H	—	—	—
D-A-M Accuracy (65)	—	H	—	H	—	L
D-A-M Nose Size (64)	H	—	—	—	—	—
D-A-W Breast Size Shift (61b)	H	—	H	H	H	—
Rorschach T/1R, Card VI (132)	—	—	H	—	—	H
Mult Choice Ror, Sex Resp (144)	—	—	—	—	H	—
Changes in Expected Direction	4	2	4	2	2	2
Changes in Opposite Direction	0	0	0	0	0	2

schach. The size of the nose, a phallic symbol, on the drawn male figure shows a tendency to increase.

By way of dynamic explanation of these results, it appears that prior to the operation the schizophrenic has regressed to a state of undifferentiated sexuality. The lobotomy apparently activates a reintegrative process whereby the patient progresses from the undifferentiated state to a level where there is some awareness and concern over sexual matters, particularly as regards the role which the patient assumes. It may be that the lobotomy reactivates conflicts related to castration anxiety and biological survival, affording the opportunity for a more realistic, less schizophrenic solution to those problems which contributed to the original development of the disease process. Kempf (19) has pointed out the importance of these issues for recovery in schizophrenia.

D. Hypothesis IV

The last of the four major hypotheses holds that there is an increase in the rate of action or behavior, or a diminution of inhibitory factors, following lobotomy. This was readily apparent in the patients' general postoperative test behavior. They were hasty and impulsive, disliked being blindfolded, and reacted negatively toward electrical contacts and repetitive, monotonous tasks.

Fourteen variables were selected by the senior author to test this hypothesis, and the direction of change was predicted. These variables, it was reasoned, were those which would best reflect alterations in the speed of action. (Rorschach cards I and II were not used as variables in this study because of the change in test conditions required by this technique. Subjects were called out individually from the group tests to be examined by a different examiner in a different room. Many subjects appeared to have difficulty with the first two Rorschach cards while making adjustment to the new situation. Cards III through X inclusive did not show this difficulty and are therefore included as variables for this hypothesis.) The results of the statistical analysis are found in Table 5. Of the 84 possible changes, 37 are observed at the .10 level or better, with 30 in the expected direction. This over-all result is presumed to be significant at the .05 level. The individual time periods are significant with the exception of the pre to two-week and

TABLE 5

DIRECTION OF NET CHANGE IN VARIABLES RELATED TO HYPOTHESIS IV: INCREASED RATE OF ACTION

(H indicates increased, L indicates decreased, — indicates no change in rate of action)

Variable	Time Period					
	Pre– 2 wk.	Pre– 3 mo.	Pre– 1 yr.	2 wk.– 3 mo.	2 wk.– 1 yr.	3 mo.– 1 yr.
Expected Change	H	H	H	H	H	H
Motor						
Downey Time for USA (35)	—	H	—	—	—	—
Downey Total Time (36)	H	H	—	L	L	L
Finger Dexterity (29)	—	H	H	H	H	—
Cognitive						
Serial Seven Time (6)	H	H	H	H	H	H
Word Assoc Time (88)	L	—	—	H	H	—
Digit Symbol Score (2)	—	H	—	H	H	—
Reaction Time, Rorschach						
Card III (129)	H	—	H	—	H	—
Card IV (130)	H	H	—	—	L	—
Card V (131)	H	—	—	—	—	H
Card VI (132)	—	—	H	—	—	H
Card VII (133)	—	L	—	—	—	—
Card VIII (134)	—	—	—	H	—	—
Card IX (135)	L	—	—	—	H	—
Card X (136)	—	—	—	—	H	—
Changes in Expected Direction	5	6	4	5	7	3
Changes in Opposite Direction	2	1	0	1	2	1

the three-month to one-year post comparisons.

Supportive evidence for this hypothesis is found in the decreased reaction times for several of the Rorschach cards, in the consistent speeding up of the time for serial sevens, in the reduction of time scores for the Word Association Test, and in the increased score on digit symbol (which is in part a function of speed). Also, there is increased speed in finger and tweezer dexterity; the time is faster for the Downey speed task; the change in the inhibition score on the Downey indicates freer action. Although the inhibition score on the Downey tends to increase following the two weeks post-testing, the one year post level is still lower than the pre score. Perhaps related to the lack of motor inhibition is the expansiveness shown on the Draw-A-Man, where there is a steady increase in the size of the drawing. These results are consistent with the theories of Cobb and the work of Robinson as presented by Freeman and Watts (9).

> While the primary concern thus far has been motor speed and inhibition, there is some reason to believe that there may be a corresponding lack of inhibition in the moral-social field. Support for this notion is found in the Aversion Test, where lobotomized patients indicate less aversive attitudes towards certain asocial activities such as theft, eating cooked human flesh, murder, masturbation, and deviant sexual behavior. Only with respect to masturbation is there an exception to this trend, towards which there is more aversive feeling from the preoperative period to the period two weeks later.

There is evidence to suggest that, while the rate of action has been somewhat altered by lobotomy, the patient's basic personality is not radically changed. It was noted, for example, that although the Rorschach reaction times were reduced, the rank order for the reaction times to the individual cards remained the same. It may be, then, that what lobotomy does is to permit freer action of the pathological prelobotomized personality core rather than effect any essential change of structure.

The authors observed, in a gross clinical way, that the patients after lobotomy apparently were not concerned with realistic plans for the future. Dr. James C. Fox, Jr., consultant in neurology and psychiatry at the Northampton VA Hospital, noting that the authors had not included in their battery a test which would reveal attitudes toward the future, suggested the Robinson-Freeman questionnaire (9). It was determined in a subsequent study which employed this technique that lobotomized patients do in fact manifest significantly less concern with the future than do control patients who are tested at the same time interval.

E. Residual Variables

The reader will recall that 145 variables are listed in the section on test materials (III, C, above). We have discussed so far the changes observed in 73 variables. Twenty-two additional variables, also showing significant changes, all pertain to the Szondi technique and will not be discussed here. Five more are functions of variables already discussed, e.g., average reaction time in Rorschach, difference between immediate and delay memory, etc. The remaining 45 variables were used in setting up hypotheses either in the longitudinal analysis or the intra-individual analysis, and were not subjected to statistical test. The reader should not infer, however, that no other variables were used in the first two phases of the analysis, for this was not the case.

V. DISCUSSION

It will be recalled that the authors cited, at the beginning of this paper, three major criticisms of the literature in the area of psychosurgery. Let us reiterate these criticisms and consider them in the light of the results of the present experiment. First, the disagreement between experimenters as to the nature of the changes in psychological tests following lobotomy seems to be, in part at least, a function of the length of interval between testings. As we have seen from the support of our first hypothesis, test scores change from one postoperative period to another. Thus experimenters who examine patients six months after operation may be expected to obtain different results from those of investigators who use the same tests at three months or one year postoperation. A second factor contributing to conflicting results is the matter of control patients. We found that our control patients varied considerably from one testing to another and these changes had to be taken into account in evaluating the changes in lobotomized patients. Studies which fail to consider the variability found in non-lobotomized psychotic patients will often arrive at erroneous conclusions as to the effect of prefrontal lobotomy.

A second major criticism raised in the literature is that important psychological variables have been neglected. The wide variety of tests used in this study and the many significant test changes observed would indicate that there are available in the psychologist's armamentarium many techniques for examining the effect of lobotomy. However, the present study unfortunately lacked measures of social dynamics. The work of Greenblatt and his associates (13) on sociometric constellations may offer a methodology for the study of such factors.

The third major criticism, dealing with the lack of congruence between clinical observations and test data, is only partially answered by the comprehensive test battery used in the present study. Although we tested a variety of functions, we were concerned primarily with the cognitive sphere, examining such attributes as attention, concentration, memory, etc. We also attempted to investigate, however, certain personality alterations of the type which have been noted clinically, such as apathy, inertia, decreased level of effort, lack of concern about the future, etc. We found support for three kinds of test changes which were related to the issues of clinical evaluation: (*a*) patients tend to react faster, (*b*) patients show some increased concern over sexual issues, and (*c*) there is an increase in social contact. Unfortunately, we are unable to determine at this point the ethical significance of these changes, although on some of our tests, such as the Aversion Scale, we noted increased acceptance of antisocial attitudes.

Furthermore, we are almost completely unable to measure such things as—to mention only two—anxiety and creativity. Both of these are all-inclusive, elusive terms whose meaning is not at all clear to psychologists or psychiatrists. We can measure related atomistic variables such as foresight, planning, tension, changes in PGR, etc., but we cannot capture the essence of the gestalten. Clinical inferences which are made concerning the presence of anxiety, for example, are usually made on the basis of case studies, whereas a particular test response connotes anxiety only when it is found in a particular matrix. For this reason, a considerable amount of inferential thinking is done, and one can never be quite certain of the reliability and validity of one's conclusions. An interview technique of some type might be used to study this important area.

Last, we would like to consider a limitation of the experimental method used in this study. This limitation is a function of the relative homogeneity and heterogeneity of the subjects. For certain variables (those more closely related to biological factors) our group is sufficiently homogeneous, but for other variables (those more psychosocial in nature) our groups are probably too heterogeneous. The changes found in this experiment that have been discussed above were changes in groups. It should not be forgotten that, in so far as personality status is concerned, there is relatively little consistency from one patient to another. It appears that the preoperative personalities of the schizophrenics are so highly individualized that they have their own peculiar mode of reaction to the operation per se. Moreover, our present qualitative methods do not appear able to assess the direction or severity of the regression so that we might understand the specific determining role of the operation.

There is still another factor which has been neglected in explaining what happens following lobotomy. This is the matter of the interaction between the personality and its social environment. We have seen that the lobotomy operation is accompanied by considerable, though temporary, organic changes. These produce alterations in the way in which the patient handles his environment. The reaction of other individuals in the environment will, in turn, alter the patient's self-concept and goals, and so will result in a new personality configuration which will then produce still different modes of behavior. Although these new modes of reaction result from a surgical assault upon the organism, they are more directly a function of the reciprocal impact of personality and environment. That is, the operation sets off a chain reaction in which the traumatic event, the operation, plays a comparatively minor role. Although we have not grappled comprehensively with this issue, our results do suggest that the interaction between the environment and the behavior of the lobotomized patient can explain a good deal of the ultimate behavior of the patient. Thus, the changes which we see at one year postoperation are not so much the results of neurological degeneration as they are the attempts of a more uninhibited personality at three months postlobotomy to make new adjustments to his environment. At one year, then, we are faced with secondary personality characteristics which are "reaction formations" to the debilitation of the two-week interval and the lack of inhibition, the extraversion, and the expansiveness of the three-month period. Only a study of changes at intervals beyond one year will give us any information about the permanency of the changes noted thus far. Prediction on the basis of our present information is almost impossible because of the individual modes of reaction and the influence of the social forces which act upon each patient differently.

VI. SUMMARY AND CONCLUSIONS

An experiment was performed to study the effects of prefrontal lobotomy on white, male schizophrenic patients. Twenty-two patients, the experimental group, underwent a standard open Lyerle-Poppen prefrontal lobotomy. The control group of 22 patients was fairly well equated for age, education, length of hospitalization, diagnosis, cooperative level, and certain behavioral characteristics. All 44 patients had been approved for lobotomy by a board of psychiatrists following an unsuccessful course of electroconvulsive shock treatment, and were receiving the same types of adjunct therapy.

All experimental subjects received an intensive battery of psychological tests *before operation, and also two weeks, three months, and one year following operation.* Control patients were tested at comparable intervals. The battery consisted of 29 major tests from which 145 variables were derived. Owing to lack of cooperation or unavailability of the subjects, not all patients received all tests or could be scored for all variables on each occasion.

There was a threefold analysis of data which allowed for successive verification of hypotheses by independent judges using different samples of the total population. In the first phase of the analysis, the senior author derived certain tentative conclusions as to the nature of the changes which occur in the reorganization of personality following lobotomy. His conclusions were based on an intensive integrated analysis of all available psychiatric and psychological data for a small sample of the lobotomized group. A less exhaustive analysis was made for the remainder of the lobotomized group.

Concurrently, independent judges carried on longitudinal analyses of major tests, utilizing the data from all experimental and control patients who were tested at least once. Here certain impressions were gathered concerning change or lack of change in variables reflecting adjustment in such areas as motor, social, etc. Those impressions which were in accord with the tentative conclusions derived from the intra-individual analysis were retained as working hypotheses. Four hypotheses were thus evolved. On the basis of these hypotheses, the senior author made predictions as to the nature of change in each of 73 variables. These predictions were then subjected to statistical test. Twenty-two variables pertaining to changes in the Szondi Test were not discussed in this paper. The remaining 50 variables include five which are linear functions of other measures and 45 which were used in setting up hypotheses and were not statistically tested.

The final stage of the threefold method of successive verification involved finding the statistical significance of the *net change* in the experimental group from one administration of tests to a later administration. The standard error of the difference between the experimental-group change and the control-group change was calculated. Changes in dichotomized data were evaluated by a formula measuring the statistical significance of the net change in proportions. Here again the standard error of the difference is based upon pre and post scores of both populations. The data used in this phase of the analysis were those obtained from patients who were tested on at least two occasions.

The conclusions reported in this paper are the end product of the threefold process—i.e., the results of statistical tests of the four hypotheses which had been evolved from the extensive case analyses and from the longitudinal analyses of test performance.

1. There is a predictable pattern of test performance following lobotomy. Performance on tests measuring mental efficiency and organicity (malfunction due to organic factors) drops sharply during the two-week postoperative period. As recovery from the operation takes place, test performance returns to the preoperative base line or rises slightly above it, reaching a peak, in our study, at three months postoperation. There then appears to be a trend toward decline, such that the performance in some tests at one year postoperation falls below the base line, although several tests remain above. It is believed that the tests which drop following the three-month peak reflect permanent organic deficit.

2. There is an increase in ego boundaries (ability to differentiate between self and non-self) following operation, as shown by better performance in tests reflecting degree of contact with the environment. There is a tendency toward increased contact immediately after lobotomy which continues throughout the three-month postoperative period. The results at one year postoperation are somewhat less clear-cut, but there appears to be a decline in test performance at this time.

3. There is an increased rate of action (with no increase in accuracy) postoperation such that lobotomized patients are able to accomplish more in a given length of time. This increased motoric speed begins shortly after operation, reaches a peak at three months postoperation, and then tends to decline. Analysis of projective tests suggests that there is no essential reorganization of the basic personality, although lobotomized patients are more impulsive, less concerned with the future, and less averse toward asocial thought and behavior.

4. There is increased awareness of sexuality following lobotomy, which seems to be maintained for at least three months As in the conclusions reported

above, the course following three months is somewhat uncertain. Dynamically, this finding seems related to anxiety over issues of castration and biological survival.

The authors recognize certain weaknesses in this study, such as the lack of measures of social dynamics, the inability of our test battery to determine the ethical significance of changes following lobotomy, and the failure to measure global variables such as anxiety or creativity. A further limitation is found in the experimental method in that individual reactions to the lobotomy, stemming from the highly individualized prelobotomy personality of each patient, are not fully considered. To some extent, however, the clinical analysis of each patient (which formed the first part of the threefold analysis) does take individual reactions into account. Related to this is the matter of interaction between the personality after lobotomy and the environment; it appears that changes at any one period may be partly determined by reactions of persons in the environment to behavior of the individual. A projected study of changes beyond one year will throw light on this matter and will also give information as to the permanency of the reported changes. This study—as well as validation, in an independent group, of our present findings—is now in progress.

APPENDIX A

The following report, excerpted from a forthcoming article (abstracted in [31]), illustrates the qualitative analysis of a single case using data from psychological tests, psychiatric observations, and case history material.

The patient is a 35-year-old, single male who completed three years of high school. He became mentally ill after three months of army service, complaining of depression, suicidal preoccupation, insecurity, and strange feelings; and was hospitalized with a diagnosis of schizophrenia. Later he alternated between periods of hyperactivity and apathetic inaction, developed delusions of syphilitic infection, and felt that his pelvis was tipped. It was also determined that he had been disturbed about masturbatory practices since age six, and was quite concerned over the suicide of an 11-year-old brother. At this hospital, he was unsuccessfully treated with insulin subshock and electroconvulsive therapy, and was lobotomized in December 1948.

His general test behavior at the four time intervals approximated that of a normal individual. He was cooperative, alert, attentive, and expended a considerable amount of effort on the tests.

Let us now consider the patient's drawings of the male human figure. Comparing the drawings before lobotomy and two weeks after, we note that on the post drawing, the hair has been shaved off, there is a large concavity at the approximate site of the operation, and that the nose has been elongated. Assuming that the observed changes are attributable to the intervening lobotomy, it may be posited that they express the patient's conscious or unconscious response to the biological assault in terms of his body image. That is, the reflection of actual body changes in the drawings—the absence of hair and the alteration of skull contour—would seem to substantiate the notion that the drawn figure is a projected conception of the self. If we assume that the nose is a phallic symbol, then the postoperative lengthening of the nose would indicate that lobotomy occasions personality changes related to psychosexual status. It is interesting to compare his drawings with a response to card IV of the Rorschach. To the same area where he had reported "a cow's head" prior to lobotomy, he responded postoperation, "Old man, big nose, eyes are tight, just about ready to get out of a deep sleep." This would seem to lend support to the hypothesis that responses to projective techniques are at least in part determined by the self-concept.

An analysis of the patient's responses to the Tendler Emotional Inquiry affords us another insight into the effect of lobotomy upon the personality. Before operation the patient characteristically expressed, in his completion of sentences, undifferentiated feelings and attitudes of a primitive, regressed nature. After operation his responses were better structured, less ambiguous, and more pertinent to the stimulus demand of the sentence fragments. His ego had seemed to become less amorphous; there was greater ego control with reduction of chaotic primitive thinking.

It will be interesting to dwell for a moment on the response of the patient to one particular item in the Aversion Scale. The item in question is: "Somebody in your immediate family (sister, brother, father, mother) asks you to have sex relations with them, you oblige." When questioned before operation about his attitude, he responded in such a way as to indicate that he would be only slightly concerned over incest. At the two weeks posttesting, however, he expressed distaste for this idea by marking the score at the extreme adverse end of the scale with many heavy rounded strokes. The motoric elaboration alerted the author to the possibility that incest fantasy or an incestual act might be playing an important role in this case. To this end a careful check of the psychiatric record, especially amytal interviews and electroconvulsive therapy postcoma verbalizations, was made, revealing the following information. At the age of six, the patient had had TB and was confined to bed. He said that all he could think of at that time was sex, and that, longing for it one day and telling his mother about it, he asked her to lie down upon the bed naked. She complied with his wish, he went on, and then he had had sexual relations with her. The fact that this peculiar response occurred only at the two-week testing suggests that the immediate effect of lobotomy is to lower the threshold for the expression of repressed material. It would appear, too, that biological assault increases the sensitivity of the organism to events in the present which are related to previous emotional trauma.

APPENDIX B

The following example of the type of information gathered in the longitudinal analysis of test performance is summarized from the report submitted by Alfred Goldman, intern clinical psychologist. Similar analyses were made for each of the 29 major tests in the battery.

ANALYSIS OF INDIVIDUAL RORSCHACHS

Limitations

The following factors limit the validity and number of inferences which may be made from the data.

A. The sequence or training effect:

Since the same technique was used repeatedly, the effect of the prior administrations upon subsequent test results must be considered. There is evidence that the memory effect was rather substantial (e.g., giving a response to card IV that had previously been given to card III in a prior administration). Aspects of test performance which are particularly sensitive to this effect are:

1. Reaction time, which may be reduced because of increased familiarity.

2. Content, which may be affected by the memory of previous responses.

B. The effect of different examiners:

1. It is recognized that a Rorschach protocol is affected to some degree by the personality of the examiner.

2. Several divergent methods of administration, inquiry, and scoring were utilized by the different examiners at the various time periods. This could have been controlled only by using the same examiner at all times.

Variables Considered and Tentative Conclusions

A. Contact, as reflected in various aspects of test performance mentioned below:

1. Although markedly increased *P* (ability to accept and express conventional modes of thought) is found in only two records, there is some slight over-all tendency to offer more conventional concepts. The mode of expression is also somewhat more conventional.

2. Style of verbalization indicates, in about 40 per cent of the records, a more tentative attitude toward percepts (reflected in the use of "might be a bat, could be a dog," etc., as compared with "is a bat, is a dog," etc.).

3. Stylistic factors as well as other types of evidence suggest that some patients drop a certain quasi-omniscient quality from their verbalization.

4. Form accuracy ($F+$) is increased in 15 per cent of the control group, with another 10 to 15 per cent tending in that direction. No such change is observed in the experimental group.

5. Only two cases show clear evidence of increased organizing ability (*Z*).

6. Sensitivity to subtler nuances in the environment (*Fc*) presents a divergent picture. Statistical treatment might indicate a bimodal distribution.

7. Although 60 per cent of the control cases indicate some increase in differentiation and/or elaboration of concepts (i.e., "animal" to "dog," "man" to "man bending down with black suit on"), none of the experimental cases show this.

B. Expression of impulse:

There seems to be more facile expression of impulse as manifested by:

1. A slight although inconsistent decrease in reaction time.

2. Increased animal movement responses (*FM*).

3. Increased control over diffuse emotionality in some records (decrease in *CF* relative to *FC*).

Other factors require consideration since they appear to contraindicate the above findings:

1. In 25 per cent of the cases there is consistent increase of *CF*, C_d, and C_n responses. This raises the question whether the increased control of emotionality is control per se, or whether the patients are merely less stimulated by color.

2. One of the most consistent findings was a decrease of morbid interest in sexuality (homo- or hetero-). This raises the question whether there is modification of impulse expression, or whether there is change in the nature of the impulse expression, or whether there is change in the nature of the impulse. There is, however, greater sex content at two weeks postlobotomy, which generally drops out by one year postoperation.

3. In three of the "improved patients" an interesting factor is observed: the emergence of an emotional superficiality that was not present before lobotomy—for example, increased *F/C* and *C/F* responses, increased Map and Medical Book content, etc. This suggests a new type of reorganization, a more successful facade.

C. Anxiety:

1. From a summary analysis of the individual Rorschachs it appears that some patients are less threatened by the environment.

2. An apparent decrease in anxiety is a concomitant of this. This is the most salient change in the experimental records.

D. Organic signs:

1. In some patients, there appears to be some awareness of assault, as reflected in increased attention to or anxiety about the head.

2. Increased concreteness.

3. Increased perseveration. Here, as in *Fc*, there is something of a bimodal distribution.

E. Flatness:

This factor, in addition to the anxiety factor, is one of the most consistent ones found. In all patients in which the affective quality of the protocols change, the direction is towards flattening. Associated with this are the following three observations:

1. One year postoperative protocols seem to be less imaginative and creative than the preoperation records.

2. There is less overt expression of aggression.

3. One year post protocols closely approximate those of the "burned out" praecox, while many of the pre records resemble protocols of excited schizophrenics.

F. Social Maturity:

There is some slight indication that the level of social maturity is somewhat higher, as manifested by an increase in constructive concepts. For example, instead of "men fighting, pulling" we find "men lifting, washing together," etc.

Conclusion: The Importance of the Unique Personality

It is assumed that the operation has psychological meaning in terms of physical assault. Just what this assault will mean, i.e., how it will be interpreted by the patient, is a function of the unique personality structure. Any statistical or pseudomathematical evaluation of Rorschach protocols may tend to mask changes associated with the physical assault, since it appears that the operation affects patients in different ways. No significant interpatient consistency was found which could be attributed to the operation.

Analysis of the data suggests that reactions to lobotomy are highly individualized and dependent largely upon the premorbid personality structure and the particular symptomatology. Thus, one patient changed in the direction of greater approximation to his potential, while another, with a similar premorbid personality, lost most of the potential that had been present, though unrealized, in the preoperative protocol. It is felt that an analysis of individual records from the point of view of the meaning of the physical assault might prove useful in understanding the way in which this experience is integrated during the reorganization of personality following lobotomy.

REFERENCES

1. Adjutant General's Office & Wechsler, D. *Wechsler Mental Abilities Scale, Form B.* Washington, D.C.: Mimeograph, 1942.
2. Babcock, H., & Levy, L. *Test and manual of directions: The revised examination for the measurement of efficiency of mental functioning.* Chicago: Stoelting, 1940.
3. Bender, Lauretta. A visual motor gestalt test and its clinical use. *Amer. Orthopsychiat. Ass., Res. Monogr.*, 1938, No. 3.
4. Benton, A. L. *A visual retention test for clinical use.* New York: Psychological Corp. 1946.
5. Capps, H. M. The vocabulary changes in mental deterioration. *Arch. Psychol.*, 1939, 34, No. 242.
6. Deri, Susan. *Introduction to the Szondi Test.* New York: Grune & Stratton, 1949.
7. Downey, June E. *Will-temperament and its testing.* Yonkers-on-Hudson: World Book Co., 1923.
8. Frank, J. Some aspects of lobotomy (prefrontal leucotomy) under psychoanalytic scrutiny. *Psychiatry*, 1950, 13, 35-42.
9. Freeman, W., & Watts, J. W. Psychosurgery: an evaluation of 200 cases over seven years.

J. ment. Sci., 1944, **90,** 532-537.

10. FULTON, J. F. *Frontal lobotomy and affective behavior.* New York: W. W. Norton, 1951.
11. FULTON, J. F., ARING, C. D., & WORTIS, S. B. (Eds.) *The frontal lobes. (Proc. Ass. Res. nerv. ment. Dis.,* 1947, **27**) Baltimore: Williams & Wilkins, 1948.
12. GOLDSTEIN, K., & SCHEERER, M. Abstract and concrete behavior. An experimental study with special tests. *Psychol. Monogr.,* 1941, **53,** No. 2 (Whole No. 239).
13. GREENBLATT, M., ARNOT, R., & SOLOMON, H. C. (Eds.) *Studies in lobotomy.* New York: Grune & Stratton, 1950.
14. HALSTEAD, W. C. *Brain and intelligence.* Chicago: Univer. of Chicago Press, 1947.
15. HARROWER-ERICKSON, M. R., & STEINER, M. *Large scale Rorschach techniques.* Springfield, Ill.: C. C Thomas, 1945.
16. HUNT, H. F. *The Hunt-Minnesota test for organic brain damage.* Minneapolis: Univer. of Minnesota Press, 1943.
17. HUTT, M. L. *A tentative guide for the administration and interpretation of the Bender-Gestalt test.* Washington, D.C.: Adjutant General's School, 1945. (Restricted)
18. KALINOWSKY, L. B., & HOCH, P. H. *Shock treatments, psychosurgery, and other somatic treatments in psychiatry.* (2nd Ed.) New York: Grune & Stratton, 1952.
19. KEMPF, E. J. Bisexual factors in curable schizophrenia. *J. abnorm. soc. Psychol.,* 1949, **44,** 414-419.
20. KING, H. E. Intellectual function. In F. Mettler (Ed.), *Selective partial ablation of the frontal cortex.* New York: Hoeber, 1949. Pp. 178-207.
21. MCNEMAR, Q. *Psychological statistics.* New York: Wiley, 1949.
22. METTLER, F. (Ed.) (Columbia Greystone Associates) *Selective partial ablation of the frontal cortex.* New York: Hoeber, 1949.
23. METTLER, F. Introduction to summary. In F. Mettler (Ed.), *Selective partial ablation of the frontal cortex.* New York: Hoeber, 1949. Pp. 475-476.
24. O'CONNOR, J. *Born that way.* Baltimore: Williams & Wilkins, 1928.
25. PORTEUS, S. D., & PETERS, H. N. Maze test validation and psychosurgery. *Genet. Psychol. Monogr.,* 1947, **36,** No. 1.
26. RAPAPORT, D., GILL, M., & SCHAFER, R. *Diagnostic psychological testing.* Chicago: Yearbook Publishers, 1945-46. 2 vols.
27. REICHARD, D., & RAPAPORT, D. The role of testing concept formation in clinical psychological work. *Bull. Menninger Clin.,* 1943, **7,** 99-105.
28. REITMAN, F. Observation on personality changes after leukotomy. *J. nerv. ment. Dis.,* 1947, **105,** 583-589.
29. RORSCHACH, H. *Psychodiagnostics.* (Trans. by P. Lemkau and B. Kronenberg) New York: Grune & Stratton, 1942.
30. ROSENZWEIG, S. The picture association method and its application in a study of reactions to frustration. *J. Pers.,* 1945, **14,** 3-23.
31. SCHERER, I. W. Validation of projective data and derivation of hypotheses from a longitudinal analysis of the test results of a leukotomized patient. *J. proj. Tech.,* 1952, **16,** 386-387. (Abstract)
32. SHIPLEY, W. C., & BURLINGAME, C. C. A convenient self-administering scale for measuring intellectual impairment in psychotics. *Amer. J. Psychiat.,* 1941, **97,** 1313-1325.
33. THURSTONE, L. L., & CHAVE, E. J. *The measurement of attitude.* Chicago: Univer. of Chicago Press, 1929.
34. WECHSLER, D. *The measurement of adult intelligence.* (3rd Ed.) Baltimore: Williams & Wilkins, 1944.
35. WELLS, F. L., & RUESCH, J. *Mental examiners handbook.* (Rev. Ed.) New York: Psychological Corp., 1945.
36. ZUBIN, J. Design of the psychologic investigation. In F. Mettler (Ed.), *Selective partial ablation of the frontal cortex.* New York: Hoeber, 1949. Pp. 173-177.
37. ZUBIN, J. Rorschach test. In F. Mettler (Ed.), *Selective partial ablation of the frontal cortex.* New York: Hoeber, 1949. Pp. 283-295.
38. *The Army individual test (AIT-1).* War Department Pamphlet No. 12-12. Washington, D.C.: U. S. Government Printing Office, 1945.

(Accepted for publication January 28, 1953.)

Vol. 67, No. 8 Whole No. 358, 1953

Psychological Monographs: General and Applied

Some Physiological Concomitants of Mental Work[1]

Edward W. Geldreich
San Diego State College

I. ORIENTATION

THE PURPOSE of the present investigation is to study certain physiological responses in the light of three problems. (*a*) Do some physiological responses change significantly under conditions of light homogeneous mental work? And, if so, what is the nature of the changes? (*b*) Do the changes in some physiological responses show any degree of concomitance with the decrement in mental work production? (*c*) Are there any patterns of internal physiological responses related to mental work products, e.g., the mental work block?

In the statement of these problems certain terms must be defined. Physiological responses include those recordable visceral and peripheral responses of respiration, heart beat rate, blood pressure, level of skin conductance, and momentary changes in such conductance (GSR). Work involving successive discrimination of randomly and serially presented colors is regarded as light and homogeneous, because five color-naming (manual) responses were repeatedly called into play by only five different colors. The work of the task is called light to distinguish it from intense mental work, such as the mental multiplication of four-digit numbers by four-digit numbers.

Implied in the concept of mental work in this study is the understanding that mental work is the result of the activity of the total organism and not the mere "emission" of a few grams of brain cells, that energy is transformed and expended, and that changes in energy distribution occur within the "mental organ system" which alter the accomplishments of the individual. The results of this study support this understanding. It is foreign to good reasoning to assume that mental activity is produced without the use of energy.

Many of the errors of early studies of the relationship between mental activity and various physiological responses and energy measurements have been due to improper definition of the mental activity, faulty control, failure to recognize the relevance of momentary conditions of excitation, and incorrect presuppositions. The results of studies of total metabolic change during mental work are equivocal. Benedict and Benedict (2) obtained a 4 per cent increase in the use of O_2 during difficult mental multiplication. The increase was said to be due to the concomitant increase in circulatory and respiratory activity. This study is perhaps the most complete and precise of the many studies, yet it is not unequivocal in its findings and interpretation.

---

[1] Based on a dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy at the University of Chicago, March, 1952.

This study, begun in 1937, grew out of a research program to determine some of the physiological correlates of the effects of mental work. The program was directed by Dr. Arthur G. Bills, now of the University of Cincinnati. Through his sponsorship and his kind, critical advice and encouragement Dr. Bills served to motivate and guide the eventual completion of the study. Much of the theoretical argument grew out of his work and writings. Later Dr. Austin H. Riesen assumed sponsorship of the thesis. Through his friendly counsel and critical examination of the study the degree of precision of the interpretation of some of the results was much enhanced. It is with sincere appreciation that the writer acknowledges the assistance granted him by both Dr. Bills and Dr. Riesen.

Implied in the conclusions of Benedict and Benedict and many of the previous investigators is the point of view of the purest of central theorists—namely, that from any recordable amount of total energy expended by a mentally active organism, one must subtract the energy due to any concomitant physiological process functioning in excess of the basal function. The futility of this point of view and methodology will be discussed later.

Permitting easier interpretation as a result of improved methods and analysis are the studies of Rounds, Schubert, and Poffenberger (25), Freeman and Giffin (17), and Darrow and Freeman (12), which certainly indicate the need for additional oxygen during mental activity. However, the work of Darrow and Freeman, and Freeman and Giffin clearly indicate the inadequacy of the total metabolic technique for the measurement of energy regimentation and transformation during mental activity or work. Freeman and Darrow (16) point out that metabolic activity in different parts of the body varies from time to time in the same individual under the same and varying conditions. An increase in activity in one organ system may be counteracted by a decrease in another organ and thus show a zero change in total metabolic measurement. Freeman and Giffin present a clear choice for those who work in the field of physiological reaction dynamics: total metabolic measurements through which neuromuscular activity is poorly reflected, or some more sensitive index such as the palmar skin conductance.

In the study of physical and mental work it is important that a reliable index of energy transformation be used. The work of Darrow and Freeman, and Freeman and Giffin has established the high reliability of skin conductance measurements as an index of neuromuscular reactivity and its relationship to oxygen consumption. The positive relationship between skin conductance changes and tonus changes has been well established (18). Sears (28) and Kirby (23) noted that skin conductance level varied concomitantly with changes in mental work.

The work of Schubert (27) and others has shown that cardiac activity is an excellent index of energy transformation, that such activity is directly related to oxygen consumption.

Those who recorded respiratory changes with total metabolic rate changes during mental work (25, 27) obtained increases in respiration rate from the mental work. Since respiration changes are subject to voluntary control, their measurement is less reliable as an indicator of energy requirements than pulse rate measurements. Other studies have shown the relationship between respiration and different behavioral states; e.g., startle, expectancy, mental work, and rest (22, 29).

A relationship between blood pressure changes and total metabolic rate changes has not been established (26). However, studies have shown a relationship between blood pressure changes and neuromuscular reactivity (14, 21). The organism is provided with equilibratory mechanisms between the vascular and the neuromuscular systems through the autonomic system (20).

The present investigation is designed to take advantage of the more recent conclusions concerning the relationship between certain secondary physiological responses and total metabolism, and to use such physiological responses as indices of energy transformation and neuromuscular reactivity. Most of the older studies failed to make use of simple statistical techniques for the evaluation of obtained means and differences; the present study is designed to apply such techniques.[2]

II. METHOD AND PROCEDURE

A. Subjects

Ten male college-student (upper division and graduate) subjects were able to complete either all of the experimental program or a sufficient amount to be used in the study. Fifteen subjects started the program but five could not finish for a variety of good reasons, such as insufficient time, working schedule change, obtaining of other employment, etc.

---

[2] Critical ratios were obtained by the use of the standard error of the difference formula 26a in McNemar, Q. *Psychological statistics.* New York: Wiley, 1949. P. 65.

B. THE MENTAL WORK TASK

The mental work task was color naming. The colors—red, yellow, green, blue, and white—were presented serially and in random order by means of a Bills' psychergometer (6). The colors were painted on a continuous, 35-mm. steel film strip. The strip contained 250 color stripes, 3/8 of an inch wide. Each color appeared in a well-lighted window and was magnified by a lens which doubled the color area. The subjects discriminated each color by manually responding to a set of correspondingly colored keys. Pressing a key automatically presented the next color. The task was to learn to respond to the perceived color by pressing the corresponding colored key as fast as skill and motivation permitted.

Each subject was practiced in responding to the colors to the point where his progress seemed to stabilize before being introduced to the experimental conditions. Each had four to seven practice sessions of a half-hour or more. Some attained a speed of about one hundred colors per minute; others were able to respond to only sixty colors per minute after the practice period. The average rate during the first five minutes of work (after the practice series) for all subjects and all days was 90.5 color discriminations per minute. It was planned that the color-naming skill become sufficiently stabilized in practice that learning effects would not counterbalance continuous work effects.

The incidence and reaction time of each color discrimination were photographically recorded.

C. THE EXPERIMENTAL PROGRAM

The following general program was planned for each subject after color-naming proficiency had been attained and after each subject had adjusted and become used to the apparatus and the laboratory situation: Day 1, mental work; Day 2, motor work; Day 3, mental work; Day 4, motor work; Day 5, sensory work; Day 6, mental work; Day 7, sensory work; Days 8 and 9, mental work under anoxia; and Day 10, mental work. Each of these experimental days consisted of recorded periods of ten minutes of prework rest, 55 minutes of work, and ten minutes of postwork rest. These rest period records will be referred to as *prerest* and *postrest,* respectively.

On the mental work days (MW condition) each subject was instructed to work as fast and accurately as he could. The mental work under anoxia will not be reported here.

The motor work period was intended as a period for the exercise of the "motor component" of the color-naming discrimination mechanism. Each subject was blindfolded and instructed to operate the keyboard keys at a rate estimated to be equal to his actual color-naming rate. The subject's keyboard responses were the only indication of the fulfillment of the condition task. This condition will be referred to as MC (for "motor component").

The sensory work period was intended as a period for the exercise of the "sensory component" of the color discrimination mechanism. The colors were presented at a set rate of 90 colors per minute. Each subject was instructed to "sense the colors passively." The subject's visual regard was used as the indication of fulfillment of the task. This condition will be referred to as SC.

Control days consisted of a 55- or 60-minute period of "rest." The subjects were told to assume a state of mental repose, to close their eyes and rest without sleeping if they could avoid sleep. This was difficult to do since they were semi-reclined in the specially made tilted chair with a back which angled 110° from the seat, and were provided with a leg rest so that the ankle was at the same level as the heart. The chair was well cushioned and the head was supported. The room was quiet except for the operation of the apparatus. Sleep was natural in this condition, so care was taken not to tell the subjects they were forbidden to sleep since such instruction would produce too great an effort to stay awake and disturb the record too much. The SC condition was also sleep-inducing to most of the subjects. The constant passive (glazy-looking attitude) "sensing" of the colors was difficult for the subjects and served to induce sleep or mere rest with eyes open. For some subjects the SC condition was a better control condition than the intended control condition.

The experimental program served more as a plan of work than as a dictated schedule to be followed. Deviations from it were not more than 25 per cent. The mental work days were spaced at least two calendar days apart. In only a few cases were experimental sessions actually scheduled on consecutive calendar days. Most of the experimental sessions were spaced two calendar days apart. Since there was such spacing between sessions and since there was no evidence of transfer of work effects evident in the mental work records, minor changes in the scheduling of days from subject to subject were regarded as not seriously jeopardizing the precision of the results. Whether attitudinal fatigue effects were transferred from session to session is not known. The only alteration in the schedule of sessions was in the occasional shift in the order of sensory and motor work days and the introduction of control sessions. Some of the subjects failed to meet the requirements of the control rest condition, even after four or five attempts and the completion of the experimental sessions. For these subjects the sensory work records were substituted for control days. It may be noted that Benedict and Benedict (2) regarded their attention and control rest conditions as equally basal. At least two *control records* were used for each subject. The records were obtained from nonscheduled sessions at the convenience of the experimenter and as laboratory conditions dictated.

All ten subjects completed, and were included in, the MW determinations—40 MW days in all.

Nine subjects made up the MC condition determinations; one subject, Le, was not included in the determinations because certain key-tapping sequences would habitually initiate in him a melodic time pattern in tapping as though he were playing the piano. The variations in tapping time disturbed his record too much. Thus there were only 18 days, instead of the planned 20, of the motor component condition.

Six subjects made up the determinations of the SC condition. One subject regularly fell asleep during this condition. Three subjects could not fulfill the control condition requirement of mental repose; their SC condition records were used as control records. Thus there were only 12 days of SC condition records.

All ten subjects were included in the control condition determinations (counting the SC condition substitutions). Two subjects had three days of control. Thus there were 22 days of control condition records.

D. RATIONALIZATION OF EXPERIMENTAL CONDITIONS

The rationale for use of the motor work (MC) and sensory work (SC) conditions grows out of historical necessity and interpretation. Therefore, it is necessary to make clear some of the assumptions that were in operation during the study. (*a*) The MC and SC conditions have been introduced to satisfy the central theorists. However, it is not claimed that the tenableness of the centralists' position can be demonstrated by this study. (*b*) The conditions of MC and SC are not "pure" conditions. Both sensory and motor activity are present in the MC and SC condition behavior. Though patterned differently, the activities overlap and hence cannot be "subtracted out" from behavior and physiological activity of the MW condition.

The MC and SC conditions are artificial and cannot lead to any conclusion concerning the centralist hypothesis.

In the treatment of results, however, the procedure of adding or subtracting the changes in the MC and SC conditions from the changes produced in the MW condition was carried through to satisfy the erroneous presuppositions of the central theorists. But the best interpretation that can be made of the changes that occur during the MC and SC conditions is one which regards the changes as being associated with different degrees of mental activity, different in degree of difficulty from that mental activity involved in the MW condition. Such an interpretation may be logically made from a central-peripheral theory of mental activity. Conclusions regarding a relationship between mental work and certain implicit physiological activity must be drawn from a comparison between the mental work period of Days 1, 3, 6, and 10 and the control days, and between the pre- and post*rest* periods and the *work* periods of each day. These changes are described both as absolute and as relative (percentage) changes from the control days level and from the pre- and postrest level.

E. THE IMPLICIT PHYSIOLOGICAL ACTIVITIES MEASURED

Heart beat rate, respiration rate, relative blood pressure change and level, volumetric change of the index finger, skin conductance level, and incidence of magnitude of galvanic skin responses were recorded and measured for all subjects during all conditions.

These activities were photographically recorded by means of the Behavior Research photopolygraph (9) originally designed by Dr. C. W. Darrow. Detailed description of the techniques of recording these responses and the shortcomings of each may be obtained from the dissertation (19).

Because of the expense of the bromide paper rolls, records were made only of the ten minutes of prerest and the ten minutes of postrest for all conditions, and the alternate five minutes of the work periods of each condition. The periods of record were minutes 1-5, 11-15, 21-25, 31-35, 41-45, and 51-55, six 5-minute periods for each work period for each condition.

Since it was anticipated on the basis of preliminary research that the duration of many of the changes in the several implicit physiological activities would be shortlived when they occur during a mental work period—presumably stimulated by various mental work changes—a short time unit should be used for all measurements of these activities. Fifteen seconds were deemed sufficient in duration to reflect most of the changes. The levels and changes in levels of activity of all recorded responses were measured for each 15 seconds during all recording periods.

III. RESULTS: PRESENTATION, TREATMENT, AND INTERPRETATION

Before presenting the data obtained from the measurement of the various physiological responses and noting whether some of the implicit physiological activities change significantly under conditions of light homogeneous mental work, the results of the mental work performance—manual color naming for 55 minutes—will be considered.

The mean number and the standard error of the mean of responses per 15 seconds for each of the 6 alternate 5-

minute work record periods, of the combined four mental work days (Days 1, 3, 6, and 10), for each of the 10 subjects and for the group were obtained. Each mean for each subject is based upon about 80 measurements (four 15-seconds in a minute, times 5 minutes, times 4 days), the exact number diminished slightly only by record faults.

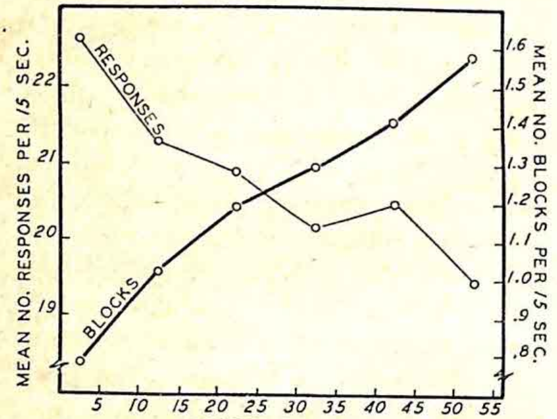


FIG. 1. Color-naming Work Curve in Terms of Average Number of Responses and Blocks per 15 Seconds for Each 5-Minute Record.

Figure 1 is a plot of the group means and describes the production trend in the color-naming performance of the 10 subjects in terms of the mean number of responses per 15 seconds for each recorded 5-minute period for the 55 minutes of mental work. As already noted, all subjects were practiced in the color-naming task.

The 55 minutes of manual color naming resulted in a decrease in each subject's rate. The decrement obtained in color-naming production is characteristic of this length of work; the production in the last five minutes was 14 per cent less than in the first five minutes. The decrease in color-naming discrimination rate is not great. In terms of number of responses per unit of time, little seemed to be lost in work rate. But it must be added that all subjects reported distinct distaste for and discomfort from the task. They were, as they claimed, tired at the end of each session. Each subject appraised the task and situation as monotonous, though none claimed to be bored. The semireclined and harnessed position may account for some of the subjects' attitude and general behavior, since it was necessary to maintain a rather fixed position to insure stable recording of the various physiological responses.

No direct measurement was made of the reaction time of each response. However, occasional indirect measurements were made by subtracting the number of blocks and total block time, and then dividing the number of nonblock responses into the total nonblock response time. There was no noticeable increase in reaction time in the regular responses. Bills (3) noted that the average reaction time of nonblock responses in color naming is fairly constant throughout an hour of work. The toll of the effect of the continuous homogeneous mental work of color naming is reflected in the increased incidence of blocks as defined by Bills (3).

The mean number and standard error of the mean of blocks per 15 seconds per 5 minutes for each of the recorded 5-minute periods, for the combined 4 days of mental work, for each subject, and for all the subjects were obtained. Again, each mean for each subject for each record period was based upon about 80 measurements. Figure 1 also provides a plot of the group means and describes the trend of increasing incidence of blocks for the 10 subjects throughout the 55 minutes of work. There are almost 100 per cent more blocks per unit of time in the last 5 minutes than in the first 5 minutes. This is characteristic of the effect of continuous simple homogeneous mental work.

The question as to whether the trend

in the response curve and the block curve is significant can be answered by an analysis of the differences between the parts of the curve. The two curves were divided into three parts; record periods 1-5 minutes and 11-15 minutes, 21-25 and 31-35, and 41-45 and 51-55 minutes respectively were combined and tested as to reliability of difference between the successive thirds and the direction of the trend. All critical ratio tests (m_{diff}/SE_{diff}) assure a better than 1 per cent confidence, except the test for the difference between the second and third thirds of the response curve. We must conclude that the possibility of chance alone determining the changes in the rates of color naming and blocking during the course of an hour of work is quite remote. The changes are due to causes found in the work operation and the organization of the responding subject. These changes receive greater meaning when noted in their relationship to some of the changes in physiological responses.

Bills (4, 5) has noted that the chief causes for blocking are the repeated excitation of the same response mechanism (the principle of homogeneity) and the tendency for different though similar response mechanisms to compete for the common efferent pathway (the principle of competition in mental work), thus resulting in faltering response in the task of color naming. The fewer the colors to respond to, the more often a response mechanism will be called into play and the greater the blocking; the more colors there are to respond to, the greater will be the competition of response outlet to a given color, and the greater will be the resulting blocking. The latter principle probably accounts for most of the decrement (4).

The question may be asked, is fatigue generated during the mental work? In terms of the work decrement of color-naming responses and the increment of blocks, we can only infer the presence of fatigue. It is generally maintained that mental fatigue is objectively reflected in its effect upon output by diminishing response rate and by the increasing momentary faltering (blocks) of the response mechanism. While this may be quite so, it does not follow that the generation of fatigue is in a one-to-one relation to the development of the product decrement. That the subjects were fatigued was evident to them. They all reported fatigue but made a distinction between fatigue and boredom. They were relieved to be finished and released from the apparatus. They all reported that their feelings did not terminate when they were finished and released. They all stretched and yawned at the end of each session. No systematic study of fatigue feelings was made. Where fatigue feelings are reported as evidence in this study, it should be recognized as evidence obtained through unsolicited fragmentary comments made by the subjects and inference from behavior occurring during and after experimental sessions. While we agree with Bartley and Chute (1) that fatigue is largely attitudinal and a function of emerging disorganization in the total individual, the implication that fatigue is so superfunctional as to be epiorganic is not accepted. Mental fatigue may be reflected in the attitudinal response of the mental worker, as well as in the disorganizational consequences of work output and the organic state.

A. Respiration

The rate of respiration will be examined first to determine if there exists a significant change as a result of mental work. Group data only will be considered in the study of this problem. The data of the individual subjects were treated in the same way as the group results and do not differ essentially. For all physio-

TABLE 1

MEAN (m) AND SE_m OF RESPIRATIONS PER 15 SECONDS FOR EACH RECORD PERIOD OF THE EXPERIMENTAL AND CONTROL CONDITIONS

Condition	Prerest 10 min.	Work Period 55 Minutes						Postrest 10 min.	Total Work	Total Rest
		1–5	11–15	21–25	31–35	41–45	51–55			
MW										
m	4.27	5.08	4.76	4.69	4.60	4.51	4.52	3.87	4.69	4.08
SE_m	.025	.046	.044	.042	.044	.048	.045	.024	.018	.018
MC										
m	4.25	4.79	4.55	4.50	4.44	4.54	4.39	3.99	4.54	4.13
SE_m	.036	.048	.051	.043	.047	.050	.051	.034	.020	.026
SC										
m	3.80	4.30	4.06	3.96	3.92	3.95	3.87	3.64	4.02	3.73
SE_m	.037	.057	.061	.064	.071	.065	.068	.039	.027	.027
Control										
m		4.13	4.00	4.07	3.90	3.91	3.88		3.98	
SE_m		.052	.044	.040	.041	.045	.045		.018	

logical records, the number of measurements for each subject for a complete day's record for the MW, MC, and SC conditions would be 40 for prerest, 120 for all six five-minute record periods, and 40 for the postrest period. Record faults and miscalculations in timing record periods diminish these numbers.

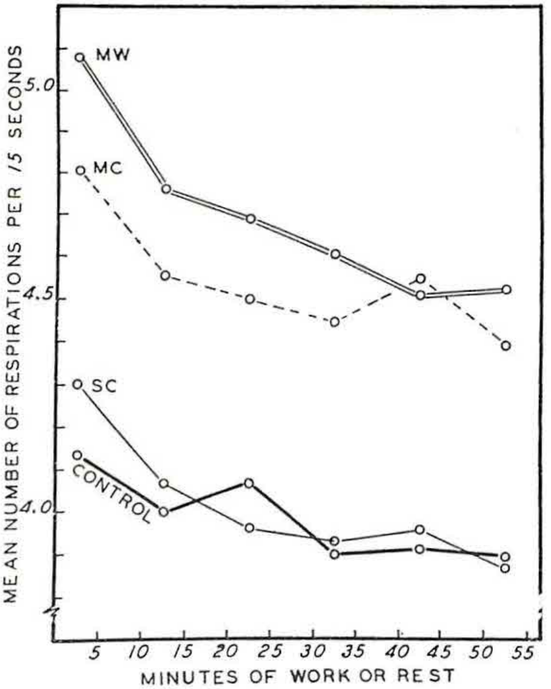


FIG. 2. Trends in Average Rate of Respiration per 15 Seconds for Successive Record Periods during MW, MC, SC Work, and the Control Rest Conditions.

Table 1 presents the mean number and the standard error of the mean of respirations per 15 seconds for each record period of the experimental conditions of MW, MC, SC, and control for the combined days of similar conditions.

To illustrate the trend in respiration rate change during the work periods of the MW, MC, and SC, and control rest, plots of the averages are presented in Fig. 2.

The curves of respiration rate change decrease for all conditions. The similarity between the control and sensory component (SC) curves is apparent. Both the MW and MC condition curves appear distinct from the control curve. The MW curve has the greatest decrement. The changes indicate a respiratory adjustment to the situation during the work and control periods. This adjustment during MW is such that respiration rate becomes less as color-naming output becomes less, and block incidence becomes greater. The decrease in respiration rate is not accompanied by an increase in respiration depth (determined by inspection), nor a decrease in the inspiration fraction. If depth increased there would be compensation for the rate decrease; likewise, if I-fraction decreased there would be an indication that the inspired air was retained longer and thus made available more oxygen absorption. But on the basis of casual examination (not reported) of all mental work records, neither of these effects occurred to compensate for the rate decrease during the 55 minutes of mental work. Note that the greatest decrease in respiration rate for all curves is during the first 15 minutes.

The SC and control curves are essentially identical in the second half. The initial high rate of respiration is a part of the behavioral mobilization and realignment to the particular task at hand. The physiological demand is greatest during the mental work task of color naming. The initial high rate of respiration during the SC condition is probably due to the abrupt need to adjust and attend to the passing colors from a level of adjustment of a sleep or quasi-sleep state. At no time during the control record periods did the average rate of respiration descend to the average rate obtained in the pre- and postrest periods of the SC condition (Table 1).

The critical ratio tests of significance of the differences between the average number of respirations per fifteen seconds during the work and during the rest periods within the same experimental condition are indicated in Table 2. All respiration rates during the work periods of the MW, MC, and SC conditions are significantly greater than the respiration rates during the prework and postwork rest periods of their respective experimental condition. For all conditions the prerest respiration rate is higher than the postrest respiration rate (see Table 1). The higher prerest respiration rate is probably due to the preparation set of the subject for the anticipated work. The lower rate of respiration during the postrest periods is the result of the release from the mobilized demands of the task. There does not seem to be any delayed payment of energy debt for the continuous work, whether it be color-naming, key-tapping, or color-sensing work, for the rate of respiration drops abruptly within the first minute of postrest and gradually continues to drop further. The resulting average rate of respiration for the postrest periods is either equal to or less than the average rate during the control rest condition (see Table 1). The tests of significance permit the conclusion that manual response color naming (MW), keyboard tapping at a self-determined rate (MC), and mere gazing at serially presented colors (SC) result in higher respiration rates than obtained during ten-minute rest periods before and after the work period.

TABLE 2

CR TESTS OF SIGNIFICANCE OF DIFFERENCE BETWEEN THE REST PERIODS AND THE WORK PERIOD WITHIN EACH EXPERIMENTAL CONDITION FOR RESPIRATION RATE

Rest Period	Work Period of Condition*		
	MW	MC	SC
Prerest	13.7	6.85	4.82
Postrest	27.3	13.8	7.87
Total rest	24.0	12.5	7.62

* Differences are in favor of the work periods.

The respiration rates for the work periods of MW, MC, and SC conditions are all significantly different. Color-naming work demanded a greater rate of respiration than mere keyboard tapping, and this, in turn, demanded more than mere sensing of the presented colors.

The tests of significance between the

control condition and the experimental conditions are shown in Table 3. All of these differences are diagrammatically represented in Fig. 3, where the average rate of respiration during the control condition serves as the base line from which are projected the absolute differences in the rate of respiration of all work and postrest period conditions. The first five-minute period of control rest was used in noting differences between the prerest period of the experimental conditions and control because of its temporal comparability. The absolute differences between the 55-minute work periods of MW, MC, and SC, and control are translated into percentages and noted on top of the respective bars in Fig. 3.

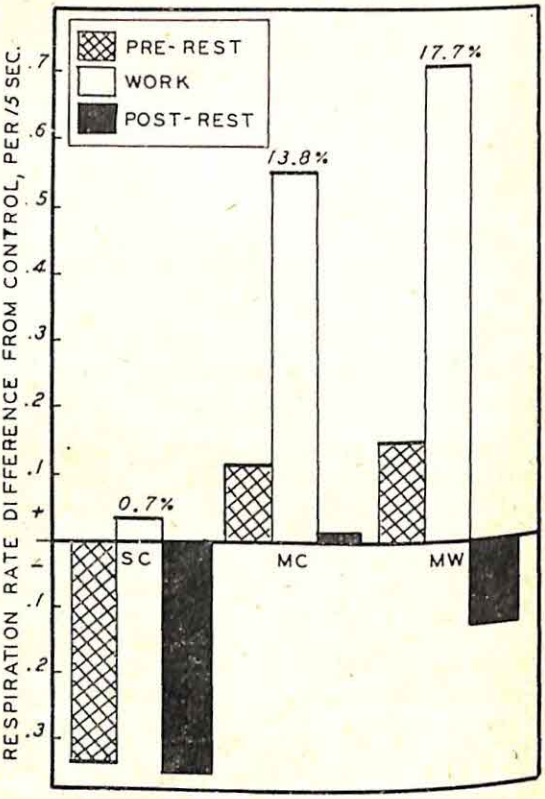


FIG. 3. Differences between the Average Respiration Rates of Control and the SC, MC, and MW Work Conditions, and between Control and Other Rest Periods.

TABLE 3

MEAN$_{diff.}$, $SE_{diff.}$, AND CR TESTS OF SIGNIFICANCE OF DIFFERENCE BETWEEN THE CONTROL CONDITION AND EACH PERIOD OF EACH EXPERIMENTAL CONDITION FOR RESPIRATION RATE

Period	CR Test between Control and Experimental Condition		
	MW	MC	SC
Prerest			
$M_{diff.}$	.142	.121	.333*
$SE_{diff.}$	.058	.064	.064
CR	2.45	1.88	5.20
Postrest			
$M_{diff.}$	.127*	.005	.341*
$SE_{diff.}$	.030	.040	.043
CR	4.19	.13	7.92
Total rest			
$M_{diff.}$	.096	.14	.258*
$SE_{diff.}$	.026	.032	.032
CR	3.77	4.42	7.95
Work			
$M_{diff.}$	.709	.551	.030
$SE_{diff.}$	.026	.027	.010
CR	26.3	20.21	2.94

* Difference is in favor of the Control condition.

The SC condition exacted only a .7 per cent increase in respiration rate over the control rest base. The pre- and postrest of the SC condition appear to be a challenge to the legitimacy of the control rest base line. Noted are the inversions; the average respiration rate is less during pre- and postrest than during control rest. If the true control rest base line were lower, it would be in favor of greater differences between it and the other conditions. But the sensory component condition is unique in its effect upon subjects. One subject, Da, could not abide by the instruction to passively observe the colors, and promptly fell asleep. Three attempts were made to get an SC condition from him. His records could not be used. All the other subjects complained of the difficulty of staying awake. The difficulty was discussed with each subject and all attempted to abide by the instructions to passively observe the colors, to maintain a fairly alert attitude, but not to fight off sleep if it overcame them, for such an effort would disturb the record. If the subject slept, the record was not used. Thus some of the subjects took advantage of the prerest condition and usually got a short nap during that period (though this was undesirable as far as the experiment was concerned it seemed to be a needed toleration), and when the postrest period came, again fell asleep. The other subjects closed their eyes and deeply relaxed during these periods. This accounts for

TABLE 4

MEAN (m) AND SE_m OF HEART BEATS PER 15 SECONDS FOR EACH RECORD PERIOD OF THE EXPERIMENTAL AND CONTROL CONDITIONS

Condition	Prerest 10 min.	Work Period 55 Minutes						Postrest 10 min.	Total Work	Total Rest
		1–5	11–15	21–25	31–35	41–45	51–55			
MW										
m	17.363	18.891	18.685	18.745	18.389	18.449	18.160	15.965	18.555	16.71
SE_m	.057	.077	.081	.087	.086	.089	.091	.057	.035	.042
MC										
m	16.811	17.017	16.615	16.456	16.428	16.452	16.387	15.682	16.563	16.264
SE_m	.094	.129	.137	.127	.116	.133	.126	.091	.052	.067
SC										
m	17.299	17.225	17.129	16.773	17.085	16.877	17.062	15.284	17.025	16.296
SE_m	.096	.124	.145	.135	.149	.143	.149	.082	.065	.071
Control										
m		16.984	16.294	16.12	16.227	15.726	15.61		16.089	
SE_m		.124	.115	.114	.109	.116	.112		.048	

the lower rate of respiration during these periods as compared with the control rest base line.

The desired attitude for the control rest condition was an attitude of mental repose, a state of wakefulness, but complete relaxation and as much freedom from mental activity as possible. This was a difficult adjustment to maintain by any alert student during the course of a school day. Perhaps it is an impossible adjustment to make! Three of the subjects encountered some difficulty in meeting the control condition requirements. One subject, Gi, fell asleep during the control rest day but was able to stay awake during the SC condition by virtue of the limited stimulation of the recurring colors. Two of the subjects, Ro and Wh, were of the "keep-busy" sort and were bored by the control rest situation and stimulated by tendencies and ideas that kept them fairly excited. They moved about too much and seemed to want to get it over with, feeling they could be doing something worth while. These three subjects had their so-called control rest day records discarded. The "work" period of the sensory component days was substituted for the control rest days. Thus, although all ten subjects make up the control group results, only six of the ten subjects make up the SC condition group results.

The inversion in the postrest following mental work is understandable. All subjects closed their eyes and deeply relaxed after the tiring 55 minutes of color naming: subjects Da, Me, Gi, and To usually slept.

The motor component condition (MC) resulted in a 13.8 per cent increase in average rate of respiration, while the mental work of color naming resulted in a 17.7 per cent increase. The tests of significance of difference (Table 3) between control rest and MW, MC, and SC work periods permit the conclusion that the rate of respiration of the experimental conditions is reliably greater than the control rest rate of respiration.

B. HEART BEAT

Table 4 presents the mean number and the standard error of the mean of heart beats per 15 seconds for each record period of the experimental conditions and of the control rest condition. These averages are the result of the combination of similar condition days and are for the group.

The six five-minute record periods for all conditions are plotted in Fig. 4 to illustrate the trend of change in average heart beat per unit time during the work periods of the experimental conditions and the control rest condition. The MW curve is notably distinct from the other curves. The trends of the MW, MC, and SC curves are similar, each tending in the direction of a reduction in heart beat rate as work progresses. The decrement in the control rest curve is greater than the curves of the experimental conditions. Using the method of testing reliability of difference between successive thirds of the curve, the decrement seems not to be due to chance (better than 1 per cent). The test of trend differences indicates that the decrease in the heart beat rate is due to factors inherent in the work

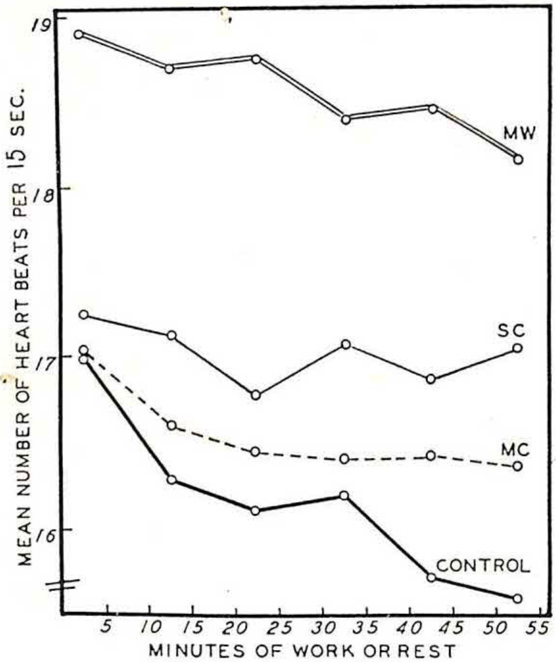


FIG. 4. Trends in Average Rate of Heart Beats per 15 Seconds for Successive Record Periods during MW, SC, MC, and Control Conditions.

tasks and operations and probably represents some progressive adjustment to the situation. The exact nature of the adjustment is a question. In the case of the SC condition, the high heart beat rate probably indicates a greater cardiac demand during this condition than during MC. The SC condition demanded a higher degree of attention for its continuance if the subjects were successful in preventing sleep than did the MC condition. The difference between the MC and SC conditions is reliable, but the difference may be due to selection. It must be remembered that the SC "population" consisted of six subjects, the MC of nine subjects. Three of the subjects who had their SC days transferred to control days were slow in their heart rate as compared to the average; one subject was a little faster. Had their records been included in the SC condition, they would tend to lower the SC condition averages.

The significance of the difference between the mean number of heart beats per 15 seconds during the work and during the rest periods within the same experimental condition is indicated in Table 5. The heart beat rate during the work period of the MW condition is quite significantly different from the rest rates before and after the mental work of color naming. All the postrest heart beat rates for all work conditions—MW, MC, and SC—are reliably less than their respective work rates; in fact, the postrest heart rates are lower than the heart beat rate during the control rest condition. But the prerest heart beat rates for the MC and SC conditions are reliably higher than the average rate during their respective work periods. Each rate during the prerest periods must be evaluated in terms of the anticipated task. It should be noted that although the MC and SC prerest heart beat rates were higher than the work rates of the same condition, the rates were lower than the prerest rate for the MW condition. None of the prerest rates was as low as the control rest condition rate (Table 4).

TABLE 5

CR TESTS OF SIGNIFICANCE OF DIFFERENCE BETWEEN THE REST PERIODS AND THE WORK PERIOD WITHIN EACH EXPERIMENTAL CONDITION FOR HEART BEAT RATE

Rest Period	Work Period of Condition		
	MW	MC	SC
Prerest	17.8	2.31*	2.38*
Postrest	38.9	8.47	16.73
Total rest	33.7	3.52	7.59

* Difference is in favor of the rest period.

Using the average heart beat rates of the total-rest periods as the base line for judgment, the conclusion is certainly permitted that color naming (MW), tapping of the keyboard (MC), and mere

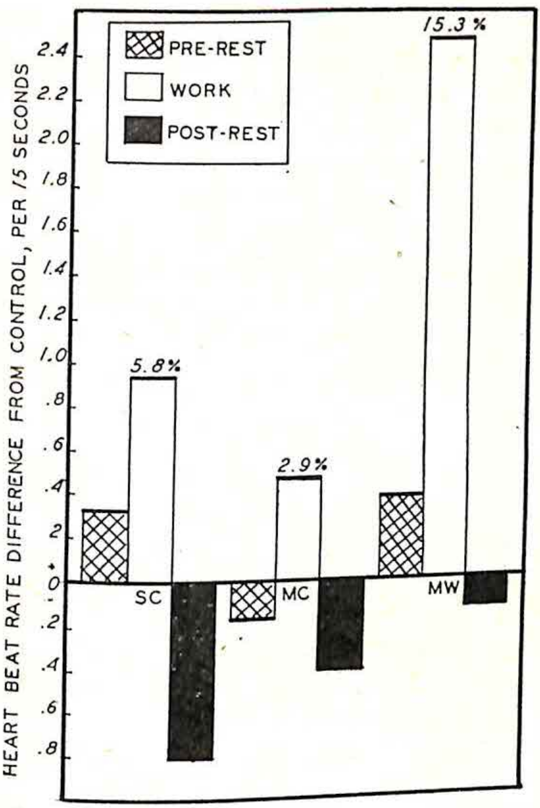


FIG. 5. Differences between the Average Heart Beat Rates of Control and the SC, MC, and MW Work, and between Control and Other Rest Periods.

looking at serially presented colors (SC) result in average heart beat rates reliably greater than the average rate obtained for rest periods during the same experimental work day setting.

The tests of significance of difference between the control condition heart beat rate and the various work and rest period rates of the experimental conditions are shown in Table 6. All of the differences listed in Table 6 are illustrated in Fig. 5 by the use of a bar diagram. The absolute differences are obtained from the base line of the average heart beat rate of total control rest for the work and postrest periods, while the prerest differences are obtained from the base line of the average of the first five minutes of control rest. The base line of these differences must be remembered when comparing the direction of differences between rest periods of different conditions in Table 6.

TABLE 6

MEAN$_{diff.}$, $SE_{diff.}$, AND CR TESTS OF SIGNIFICANCE OF DIFFERENCE BETWEEN THE CONTROL CONDITION AND EACH PERIOD OF EACH EXPERIMENTAL CONDITION FOR HEART BEAT RATE

Period	CR Test between Control and Experimental Condition		
	MW	MC	SC
Prerest			
$M_{diff.}$	.379	.173*	.315
$SE_{diff.}$	.136	.155	.156
CR	2.78	1.11	2.01
Postrest			
$M_{diff.}$	.124*	.407*	.805*
$SE_{diff.}$	.074	.102	.095
CR	1.67	3.98	8.46
Total rest			
$M_{diff.}$	.631	.176	.206
$SE_{diff.}$	.064	.082	.086
CR	9.88	2.13	2.40
Work			
$M_{diff.}$	2.466	.474	.936
$SE_{diff.}$	.059	.071	.080
CR	41.5	6.69	11.7

* Difference is in favor of the control condition.

C. RELATIVE BLOOD PRESSURE

The physiologic response record most difficult to obtain and maintain was the trace of relative blood pressure change.

Each measurement of BP trace level was made from an arbitrary "zero"—90 millimeters on a millimeter scale—at the beginning of each five-minute record period and at the beginning of the first and sixth minutes of pre- and postrest. Thus, an average value for such a record period is a statement of the level to which the record trace descended or ascended on the average from the initial level of 90 millimeters on the scale used to measure the displacement of the trace on the record.

The treatment of these measurements

TABLE 7

MEAN (m) AND SE_m OF MEASUREMENTS OF DISPLACEMENT OF LEVEL OF BLOOD PRESSURE PER 15 SECONDS FOR EACH RECORD PERIOD OF THE EXPERIMENTAL AND CONTROL CONDITIONS

Condition	Prerest 10 min.	Work Period 55 Minutes						Postrest 10 min.	Total Work	Total Rest
		1–5	11–15	21–25	31–35	41–45	51–55			
MW										
m	86.860	90.523	90.611	89.435	92.00	90.867	92.272	89.775	90.951	88.239
SE_m	.253	.245	.310	.336	.338	.350	.549	.224	.156	.152
MC										
m	86.953	88.469	89.266	88.028	90.828	89.807	90.852	87.388	89.523	87.164
SE_m	.273	.301	.484	.332	.506	.352	.416	.283	.169	.245
SC										
m	87.448	86.417	90.021	89.269	90.194	90.614	89.857	89.076	89.383	88.211
SE_m	.297	.403	.539	.366	.373	.603	.315	.313	.189	.218
Control										
m		86.575	86.423	86.494	88.563	88.329	85.676		87.002	
SE_m		.353	.276	.314	.303	.369	.333		.135	

will provide an adequate statement of the average amount of displacement of BP level from condition to condition and between successive five-minute work record periods, or rest periods, within a condition, if it is assumed that errors of measurement are approximately equal in all periods compared. The BP measurements were treated in the same manner as the respiration and heart beat rate changes.

Table 7 presents the mean and SE_m of displacements of level of BP trace per 15 seconds for each of the record periods for each experimental condition and for control. The data are the result of combining similar condition days.

Since the changes in actual blood pressure are continuous and the direction of change is relative to the previous moment's level and excitation background, the data presented in Table 7 may be translated into a form which will picture the progress of pressure change and direction. This is done in Fig. 6 with certain assumptions. One may assume that the various measured displacements in BP trace are not only representative of blood pressure change in the record periods themselves, but also representative of blood pressure changes in adjacent unrecorded periods. Furthermore, changes that do occur at one moment become conditions that partly determine the changes in the next moment. Thus if the *differences* between the base measurement of 90 made at the beginning of each 5 minutes and the average displaced level recorded for each 5 minutes are cumulated, one may plot these cumulations and thereby gain some idea of the

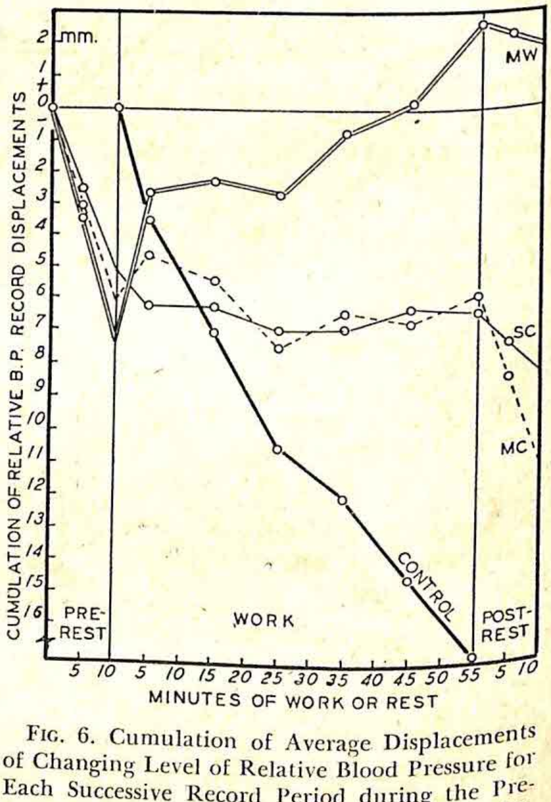


FIG. 6. Cumulation of Average Displacements of Changing Level of Relative Blood Pressure for Each Successive Record Period during the Prerest, Work, and Postrest Periods of MW, MC, and SC Conditions and during Control.

continuous change of the BP trace. The curves in Fig. 6 are not to be regarded as accurate descriptions of either relative changes in blood pressure or actual changes in the record trace of relative blood pressure. The cumulations are from averages of displacements in BP trace, not from levels of traces at the very last moment of a record period, and certainly not from that level of pressure at the very moment (an unrecorded moment) just before a record period.

The control condition resulted in the greatest change in BP trace displacement. The curve shows a rather abrupt drop in pressure during the first half of the period and lessening in the amount of decrease as the hour continues. The considerable decrease in pressure that occurred at the very end of the rest period is probably due to sleep, or to impending sleep. The changes in pressure during control are "smoother" than in the other conditions.

Figure 6 shows each curve beginning at the arbitrary "zero" value of 90. The points of the curves for the work conditions and the control condition are plotted at the end of each record period on the basis of the assumptions made. This makes for less distortion in the curve. The plots for the pre- and postrest periods of the experimental work conditions are made in the middle of these periods. The plot points for the pre- and postrest periods are the average of the BP trace displacements of that rest period. In interpreting the curves of the changes during and from prerest to the work periods in the MW, MC, and SC conditions, one is aided by examining the individual polygraph records of the changes in relative pressure. This examination reveals that the curves in Fig. 6 describe in general the course of pressure changes as they actually were recorded. During the MC and SC conditions, the pressure decreases and tends to continue so through the prerest into the work periods. During the mental work condition, relative blood pressure actually increases markedly with the onset of mental work from the decreasing lower level which occurs during the MW prerest period.

On the basis of our attempts to approximate the actual pressure change represented by the relative blood pressure trace changes, it may be conservatively estimated that the repose permitted in the control condition results in a progressive diastolic pressure decrease to about 15 mm. of Hg at the end of the rest hour. Color naming by means of manual response results in a gradual increase in diastolic pressure amounting to about 8 mm. of Hg at the end of the work hour.

TABLE 8

MEAN$_{diff.}$, $SE_{diff.}$, AND CR TESTS OF SIGNIFICANCE OF DIFFERENCE BETWEEN THE CONTROL CONDITION AND EACH PERIOD OF EACH EXPERIMENTAL CONDITION FOR DISPLACEMENT OF RELATIVE BLOOD PRESSURE TRACE*

Period	CR Test between Control and Experimental Condition		
	MW	MC	SC
Prerest			
$M_{diff.}$	.285	.378	.873
$SE_{diff.}$	.434	.446	.461
CR	.655	.846	1.89
Postrest			
$M_{diff.}$	2.773	.386	2.074
$SE_{diff.}$	.261	.264	.340
CR	10.62	1.46	6.09
Total rest			
$M_{diff.}$	1.237	.161	1.209
$SE_{diff.}$	.203	.224	.256
CR	6.09	.720	4.72
Work			
M_{diff}	3.949	2.521	2.381
$SE_{diff.}$	.206	.216	.232
CR	19.41	11.66	10.26

* All differences are in favor of the experimental conditions.

Table 8 presents the critical ratio tests of significance of difference between the control rest condition and each period of each experimental condition for the average level of BP trace. All of the differences tabulated in Table 8 are illustrated by means of a bar diagram in Fig 7. All the work period (total work) averages of the BP trace level are quite significantly higher than the average BP trace level during control rest. The pre-

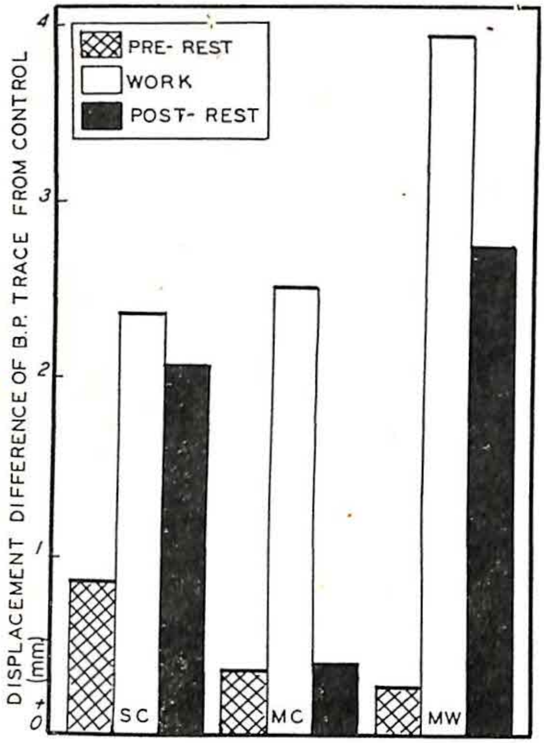


FIG. 7. Differences between the Averages of the Relative Blood Pressure Trace Displacements of the Control and the SC, MC, and MW Work Conditions, and between Control and Other Rest Periods.

rest averages of the level of BP trace for all experimental conditions are not significantly different from the control rest level. This is also true of the postrest average level following the MC work condition. The average level of the BP trace during the postrest periods of the MW and SC conditions is significantly higher than the control rest level.

Since the measurements of the actual blood pressure changes in the present study are relative, and in view of the method used to express these changes, a statement of percentage of change from a control level is improper and meaningless. We are warranted in saying, however, that mental work of the type engaged in and for the period studied does result in a definite increase in blood pressure, an increase distinct from the changes (a drop) in pressure resulting from tapping the psychergometer keys in the MC condition, or "sensing" the serial presentation of colors in the SC condition. The work of Darling and Darrow (8) suggests that during the mental work condition the autonomic balance was dominated by sympathetic innervation, while during rest, by parasympathetic innervation. During the MC and SC conditions, the balance seems to be in favor of parasympathetic innervation.

D. SKIN CONDUCTANCE LEVEL

The skin conductance level was noted continuously throughout the entire session for all conditions and phases of conditions, but was recorded photographically only during those periods already reported for the other physiological responses. The conductance level was measured directly from the calibrated bromide record in terms of units of log. micromhos × 100. Table 9 presents the essential data for the group. The mean and SE_m of conductance level per 15 seconds for each record period for each experimental condition of MW, MC, and SC, and control rest are tabulated in Table 9.

The trends in the conductance level during the 55 minutes of work of the MW, MC, and SC experimental conditions and during control rest are graphed in Fig. 8 by plotting the average conductance level of each succeeding 5-minute record period. All conditions result in a gradual decrease in skin conductance. The greatest reduction in conductance occurs during the first 15 minutes of control rest after which the trends tend to parallel the trends in the other conditions. The test of difference between the first and second thirds of the control curve indicates a significant difference

TABLE 9

MEAN (m) AND SE_m OF MEASUREMENTS OF LOG. MICROMHOS × 100 PER 15 SECONDS FOR EACH RECORD PERIOD OF THE EXPERIMENTAL AND CONTROL CONDITIONS

Condition	Prerest 10 min.	Work Period 55 Minutes						Postrest 10 min.	Total Work	Total Rest
		1–5	11–15	21–25	31–35	41–45	51–55			
MW										
m	144.738	151.068	148.988	148.638	147.150	145.02	144.144	139.734	147.492	142.44
SE_m	.602	.874	.844	.867	.857	.875	.863	.646	.344	.442
MC										
m	145.644	144.833	140.983	139.2	138.550	136.264	136.416	135.630	139.429	140.741
SE_m	1.081	1.356	1.362	1.362	1.359	1.469	1.359	1.080	.566	.773
SC										
m	149.687	146.650	142.375	143.369	141.125	141.120	140.718	132.439	142.594	141.597
SE_m	.994	1.398	1.506	1.381	1.448	1.165	1.558	1.014	.601	.766
Control										
m		138.144	131.021	129.430	128.791	126.913	127.881		130.391	
SE_m		1.331	1.296	1.272	1.189	1.244	1.284		.528	

with greater than 1 per cent confidence. The test between the second and third thirds suggests that the trend of conductance is definitely leveling off; the difference between these thirds is insignificant.

There was a strong tendency to sleep during the control rest condition. This tendency was resisted by the attitude of each subject to cooperate as best he could in maintaining a wakeful state of "mental repose." In healthy subjects the skin conductance decreases considerably during sleep. Conductance levels equal to 90 to 100 log. units (125,000 to 130,000 ohms resistance) were not unusual for those subjects who actually slept during the control rest days early in the experimental program. These control records were not used. Most of the subjects did sleep during the last fifteen minutes of control rest; at least, their condition could be called near-sleep. But despite their proneness to sleep, or actual sleep, the conductance level did not descend to the usual sleep conductance level. Perhaps the maintained level is representative of a certain degree of alertness and tension which the subjects continued to express in response to the control condition instructions even though they nearly slept.

The rough tests of trend changes in conductance level during mental work (MW) indicate a reliable decrement from third to third; at least, a 5 per cent confidence inheres in the difference between the succeeding thirds.

Table 10 presents the critical ratio tests of significance of difference between the work period and the pre- and postrest periods within each experimental condition for the palmar skin conductance. The differences are expressed in terms of log. conductance (micromhos × 100) units. The 55 minutes of mental

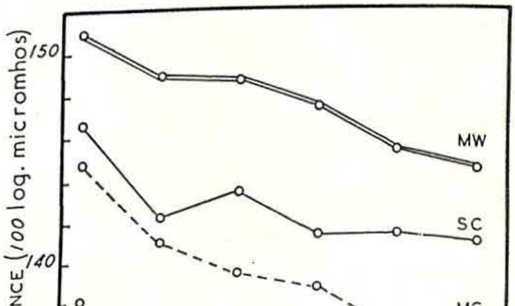


FIG. 8. Trends in the Average Palmar Skin Conductance per 15 Seconds for Successive Record Periods during the MW, SC, MC Work, and Control Rest Conditions.

TABLE 10

CR Tests of Significance of Difference between the Rest Periods and the Work Period within Each Experimental Condition for Palmar Conductance (Log. Micromhos × 100 Units)

Rest Period	Work Period of Condition		
	MW	MC	SC
Prerest			
$M_{diff.}$	2.754	6.215	7.093
CR	3.98	5.09*	6.09*
Postrest			
$M_{diff.}$	7.758	3.80	10.156
CR	10.61	3.14	8.53
Total rest			
$M_{diff.}$	5.052	1.312	.997
CR	9.04	1.37	1.02

* Difference is in favor of the rest period.

work (MW) resulted in a significantly higher average conductance level than the average conductance level in either the prerest or postrest of the MW condition. This is not true for the MC and SC conditions. In both of these conditions the average level of conductance during the prerest period is significantly higher than the average level of conductance during their respective work periods. As shown in Table 9 the trend of the average level of conductance for both the MC and SC conditions decreases from the prerest level, throughout the work periods (Fig. 8), to the lowest level of conductance during the postrest period.

The average levels of conductance for the 55 minutes of work for each of the experimental conditions are significantly different. The differences have already been noted in Fig. 8. If palmar skin conductance is an indirect measure of alertness (11, 12) and bodily tension (18) it may be noted that color-naming work required a greater degree of alertness and general bodily tension than mere passive perception of serially presented colors and mere responding by pressing keys at a voluntarily controlled rate while blindfolded. Similarly, the passive perception demanded greater alertness and tension than the key-tapping condition.

The conductance levels of the rest and work periods of the experimental conditions are higher than the control rest level. The differences and their tests of significance are noted in Table 11 where it can be seen that the only exception to a significant difference is the postrest of SC and control. These differences between control rest as a base and the work and rest periods of the experimental conditions are diagrammed in Fig. 9.

The conductance level during prerest is probably determined by the adjustment which the subject makes to the anticipated task. The conductance level represents the tension of set expectancy. The conductance levels of the prerest periods of MC and MW are essentially equal but the SC prerest level is reliably higher than the MC and MW prerest levels. The conductance level of the postrest periods is a function of the previous task. Greatest relative reduction of conductance is made after the work of sensing or gazing at the serially presented colors of the psychergometer.

The results justify the conclusions that light homogeneous mental work—manual discrimination of randomly and serially presented colors—is accompanied by an initial increase in skin conductance from a prerest level of conductance. The average initial increase in skin conductance during the first five minutes of the mental work is about 6 log. micromhos × 100 units greater than the average skin conductance obtained during a ten-minute prerest period. The average level of skin conductance during 55 minutes of manual color naming is about 17 log. micromhos × 100 units greater than the average level of skin conductance during a comparable period of mental repose. Similarly the task of key tapping resulted in an average increase in skin conductance of about 9 log. conductance units,

TABLE 11

MEAN$_{diff.}$ AND CR TESTS OF SIGNIFICANCE OF DIFFERENCE BETWEEN THE CONTROL CONDITION AND EACH PERIOD OF EACH EXPERIMENTAL CONDITION FOR PALMAR CONDUCTANCE (LOG. MICROMHOS × 100 UNITS)*

Period	CR Test between Control and Experimental Condition		
	MW	MC	SC
Prerest			
$M_{diff.}$	6.594	7.500	11.543
CR	4.51	4.38	6.95
Postrest			
$M_{diff.}$	9.343	5.239	2.047
CR	11.20	4.36	1.81
Total rest			
$M_{diff.}$	12.049	10.349	11.206
CR	17.51	11.05	12.04
Work			
$M_{diff.}$	17.101	9.038	12.203
CR	27.1	11.6	15.2

* All differences are in favor of the experimental conditions.

and the task of mere gazing at serially presented colors resulted in an increase of about 12 log. conductance units above the average skin conductance obtained during a comparable period of mental repose.

The Centralist Position

If it is assumed that the mental activity of color naming, as represented in the MW condition, can be fractionated into motor, sensory, and central components, and that the noncentral components can be subtracted out in order to determine the dynamics of "pure" mental activity, then it may be reasoned (as implied in the centralist position of Benedict and Benedict [2]) that some measure of the energy of mental activity involved in the MW condition may be obtained by suitably subtracting out the energy needed in the expression of the MC and SC conditions. It must be further assumed that the MC and SC conditions exercised completely and exclusively the motor and sensory components, respectively.

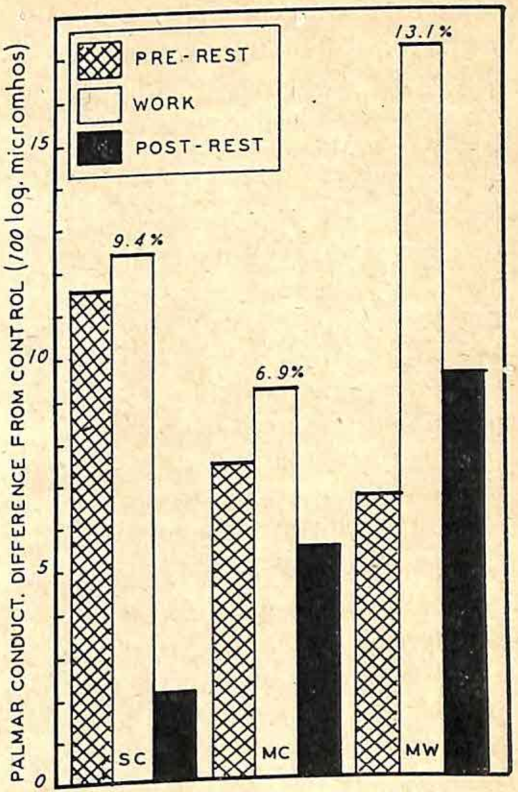


FIG. 9. Differences between the Average Palmar Conductance Levels (Log. Micromhos × 100) of Control and the SC, MC, and MW Work Conditions, and between Control and Other Rest Periods.

Of the various physiological reactions observed, heart beat and respiration rates are the best indices of total metabolism. Following through the procedure, summing the differences between control and MC and SC conditions, and subtracting that sum from the difference between the MW and control conditions, we obtain the following "residues" for respiration (Table 3) and heart beat (Table 6): 0.128 respiration and 1.056 heart beats per 15 seconds. If the centralist hypothesis of "pure" mental work and the assumptions are correct, then it might be concluded that the light, homogeneous mental work, the work performed (presumably) by the central factor, required a little more than a half respiration and four heart beats per minute. The Benedicts concluded that their 4 per cent increase in rate of metabolism for rather difficult work could not be due solely to the activity of the brain cells (as they hoped to be able to show), but most likely to the muscular activity of increased heart action and ventilation. But if the data reported in the present study are correct, and the assumptions and reasoning are valid, then it might be concluded that the half respiration and four heart beats per minute are required for both the cerebral metabolic activity *for* mental work and the supporting secondary physiological activities (respiratory and cardiac activities themselves included) of the cerebral activity.

E. Relation between Changes in Mental Work Products and Changes in Physiological Activity

The question now arises as to the relationship between the *changes* in the several physiological activities and the mental work product *changes*. The group results indicate that mental work produces significant increases in heart beat and respiration rate, blood pressure, and skin conductance.

Since each subject's results were treated in the same manner reported for the group, data are available for correlating the physiological activities measured with the mental work products. The average level of the physiological activities for each five-minute record—six such alternate five-minute record periods in the 55 minutes of mental work—were ranked and related to the ranked mental work products of average responses and average blocks for each five-minute record period. Eight rank-difference correlations were obtained for the results of each mental work day for each subject, 320 in all. The individual correlations themselves demand a very high value, 4 degrees of freedom with only 6 pairs of measurements, to be regarded as significant. But the combining of four correlations, four days for each subject, and the combining of these for all subjects reduce the magnitude of the value demanding significance. The combined correlations provide information on the presence and direction of an existing relationship between some physiological activity and mental work. We are not aware of any reported correlations between the same physiological activities and mental activities made from *coincident measurements* and therefore have no basis for comparing the magnitude of correlations obtained in the present study with others. However, it might be noted that Darrow and Heath (13) obtained correlations of the order of .30 between similar physiological activities in response to discrete stimuli and certain personality factors as measured by the questionnaire technique. Wenger (30) obtained correlations of the order of .25 between some similar physiological activities and rate of recitation, speech, counting, tapping, and tonguing.

The Spearman method of rank-difference correlation was used. First-place rank was given to that measurement having the greatest numerical magnitude in each set of six measurements of each of the two variables correlated. No correction was made for trend.

In combining the correlations for the four MW days the *z* transformation was used. The *z* transformation was also used to obtain the combined correlation for all subjects. No correction was made for similar rank values. The rho's were not translated into Pearsonian *r*'s.

Table 12 presents the combined rho's

TABLE 12

COMBINED CORRELATIONS OF FOUR DAYS OF MENTAL WORK FOR EACH SUBJECT ($df=12$) AND FOR ALL SUBJECTS ($df=30$) BETWEEN THE CHANGING MEANS OF THE WORK PRODUCTS AND THE CHANGING MEANS OF THE VARIOUS PHYSIOLOGICAL RESPONSES

Measurements Related	Subjects										Group
	Da	Gi	Gu	Ha	Le	Me	Ro	To	Wa	Wh	
Responses—Respira.	.58	.43	.65	.81	.33	.81	.59	.71	.54	.47	.62
Responses—Heart B.	.33	−.50	.86	.09	.02	.64	.69	.76	.19	.53	.43
Responses—Blood P.	−.05	−.28	.19	.25	.05	−.07	−.32	−.27	.07	−.51	−.08
Responses—Log. Cond.	.34	.77	−.09	−.04	−.03	.76	.47	.64	.32	.14	.39
Blocks—Respiration	−.41	−.13	−.57	−.55	−.46	−.57	−.39	−.32	−.92	−.58	−.54
Blocks—Heart Beat	.20	.10	−.77	−.09	−.19	−.55	−.26	−.94	−.66	−.35	−.46
Blocks—Blood Pres.	.27	−.06	−.08	−.19	.34	−.14	.21	.23	−.12	.08	.02
Blocks—Log. Cond.	−.38	−.76	−.41	.36	−.21	−.41	−.47	−.97	−.93	−.53	−.60

for each subject for each of the eight correlations for the combined four days of mental work. Each combined correlation for each subject has 12 degrees of freedom. The standard error of z for 12 df is $1/\sqrt{12} = .288$. Thus significance at the 5 per cent and 1 per cent levels would require correlations of .51 and .63, respectively. Many of the correlations approximate these levels of significance.

Table 12 also presents in the last column the combined rho's for the ten subjects. These combined correlations have 30 df on the assumption that the physiological and mental work activities of each of the ten subjects are independent random samples of the same population of such activities in human behavior. The 120 df which the combining of 40 correlations would warrant is not justified in the present combination of "four subject days" with "subjects," so the more rigorous 30 df was imposed to satisfy partially our somewhat questionable practice of combining 40 correlations. Realizing the initial source of the rho's (six pairs of measurements), the nature of the averages that are correlated, and the larger sampling error of z for rho values, we are not intending to extract precision of measured relationship; we merely wish to indicate the possible presence of relation and its direction. The 30 df provide a SE_z of .1825 which roughly suggests 5 per cent and 1 per cent confidence in rho values of the magnitude of .35 and .45, respectively.

In Table 12 the correlations between the means of the work products and the physiological responses show considerable consistency of direction from subject to subject. Of the eight combined correlations for the group, two are zero in value: the relationships between the means of relative blood pressure trace displacements and responses, and between blocks. Since the measurement of relative blood pressure is questionable, these values are to be expected.

The obtained combined correlations of .62 for respirations and responses and −.54 for respirations and blocks suggest that these work products are a function of the changing organization of the momentarily mobilized energy which is indirectly expressed through respiration, and the adjustment made by the subject to the present work situation. It may be that the increasing blocking and decreasing response rate are a function of the subject's inability to maintain the initially higher energy mobilization, which in turn may be a function of another correlate of the total alignments of the body in the performance of the simple mental work task. This correlate may be general body tension with its related changes in skin conductance. Tension is known to increase at the onset of work and to be maintained or even increased for a time, dependent upon the task set, and then to decrease to a level somewhat above that of rest (7, 15). Thus, the mental work changes may be due to the reduced respiration rate, which may produce a "relative" anoxia in the neuromuscular mechanism which is already taxed by repeated excitation; or another correlate (not known) may effect tension changes during the mental work which in turn may reduce the energy demands of the body and thus effect a reduction in respiration and also reduce the essential proprioceptive afferent excitation of the cortex (and the wakefulness center), resulting in lower mental work efficiency (and tendency to sleep).

Heart beat rate changes are likewise significantly related to the mental work products, as indicated by combined rho's .43 and −.46 for the changes in responses and blocks respectively. As the mean rate of color naming decreases and the mean incidence of blocks increases, the heart beat rate mean decreases during the mental work period. As with respiration, heart beat rate is an indirect index of total metabolic activity. Thus as the performance output of the practiced subject decreased during the working hour, the momentary energy mobilization likewise decreased. What has been said concerning respiration may be equally applied to the heart beat rate relationships.

The combined correlation of .39 between log. conductance of the skin and rate of color naming is significant at about the 3 per cent level. Since palmar skin conductance is an indirect measure of the level of alertness or apprehension of the subject, it may be said that as alertness decreases the rate of response decreases. Likewise, since conductance level is directly related to the amount of general body tension or tonus, it may be said that as general body tension decreases there occurs a decrease in the rate of color naming. Similarly, since the measurement of the skin conductance level is an indirect (and more reliable) index of varying metabolic activity (17), then as the momentary requirements of energy decrease, the rate of color naming decreases during the work hour.

The combined correlation of −.60 between average number of blocks and the average level of skin conductance is significant beyond the 1 per cent confidence level. The correlation implies that as the average number of blocks per record period increases, the palmar skin conductance decreases. Likewise, as the incidence of blocking increases the general bodily tension, mental alertness, and momentary energy requirements decrease, as measured indirectly by the skin conductance level decreases.

Skin conductance changes are a function of changes in the excitation background and susceptibility of the subject to stimulation. Stimulation results in momentary increases in conductance (GSR's) which recovers to the initial conductance level in varying amounts in about 15 seconds. The degree of recovery is a function of the altered excitation background as a result of stimulation. As will be pointed out later, blocks are associated with galvanic skin responses, and the number of GSR's decreases as the mental work throughout the hour continues. The general conductance level is less disturbed as the work hour progresses and tends to diminish. As the number of blocks increases, the number of GSR's decreases. With diminishing alertness, reduction in supporting tonus, the effectiveness of the individual color stimulus is reduced and tends less and less to result in momentary increases in skin conductance. Even the blocks cause fewer GSR's, as will be shown. This is Robinson's fifth principle of work decrement in operation (24). Robinson maintains that the work decrement of a given response mechanism is a function of the "qualitative integrity of the stimulus throughout the work period" during which changes in efficiency develop. The qualitative integrity of the colors is a function not merely of the practiced response of pressing the correct key, but of the entire adjustment which the worker makes to the total work situation and task. Changes within the worker, the work situation, and the action of work upon itself result in changes in the qualitative integrity of each of the five colors in eliciting the correct response.

All the physiological activities observed are related to the energy requirements of the mentally working subjects. In terms of the measured changes in respiration, heart beat, and skin conductance, the momentary energy requirements change during the mental work performance. The changing rate of respiration and heart beat (and relative blood pressure changes) are indicative of regulative changes in the circulatory system. The changes in skin conductance, rate of color naming, and incidence of mental blocking are indicative of changes in the neuromuscular system. The various correlations obtained are indicative of the integration existing between the two systems. The syndrome of physiological activities is a picture of progressive adjustment, of changes of internal regulation with respect to the mental work task and task effects upon the working organism.

F. Relation between Blocks and the Galvanic Skin Response

As blocking increases during the course of an hour of continuous simple homogeneous mental work, rate of respiration, heart beat, and palmar skin conductance decrease, while relative blood pressure tends to increase. All of these changing physiological activities are occasioned at levels greater than a resting level. The relationship between blocks and galvanic skin responses has already been suggested. A color-naming blocked response is characterized by a latency of at least twice the median reaction time of responses occurring within the immediate minute. This is in agreement with Bills' interpretation (3).

Figure 10 presents the plots of the number of blocks and galvanic skin responses occurring per minute for each of the six alternate five-minute record periods of the 55 minutes of color-naming work for the group. Almost a 100 per cent increase occurs in block incidence during the work period. The number of

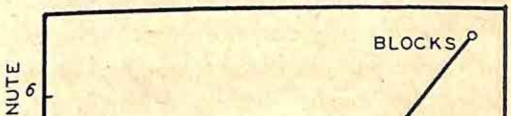


Fig. 10. Changing Average Number of Blocks and GSR's per Minute during Color-Naming Work.

GSR's per minute tends to decrease slightly during the course of the work; the decrease amounts to about 18 per cent. The last half of the mental work period shows practically no change in number of GSR's. These changes in the frequency of blocks and GSR's during the mental work must be considered in relation to the processes of block and GSR association. An attempt is made here to approach the problem of causal relationship in terms of "coincidence." The temporal definition of coincidence as here used places the problem on a broad basis. Later, an attempt will be made to describe the "causal" relation between blocks and GSR's more specifically.

The graphs of Fig. 11 describe the relationship between the blocks and GSR's in terms of the percentage of coincidence between the two phenomena. To understand further the plots of coincidence it must be recognized that there are different numbers of GSR's and blocks, and that several blocks may coincide with a single GSR, and more than one galvanic response may coincide with one long block. Thus it was necessary to make more than one notation of the degree of coincidence of block and galvanic phenomena.

Figure 11 shows the relative *decrease* in the percentage of *blocks associated with GSR's,* a decrease from 62 to 52 per cent from the first to the last five-minute record periods. However, in view of the considerable absolute increase in the number of blocks during work, the absolute number of blocks associated with GSR's increases. The same may be noted with respect to the blocks *not* associated with GSR's: an increase in absolute number and relative number (inverse of the curve of blocks with GSR's, Fig. 11).

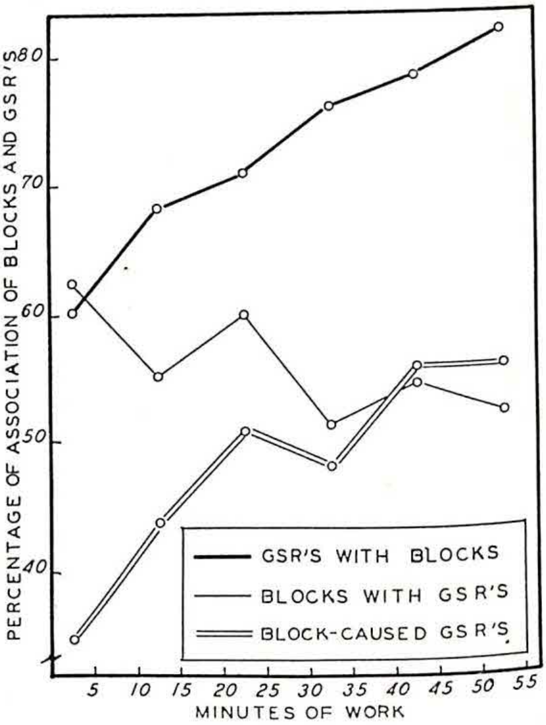


FIG. 11. Changing Percentages of Associations of Blocks and GSR's, and GSR's "Caused" by Blocks during Color-Naming Work.

While the absolute number of GSR's decreases (Fig. 10), the percentage of *GSR's with blocks* increases (Fig. 11). The relative increase is from 61 to 82 per cent from the first to the last five-minute record periods. The GSR's not associated with blocks behave the opposite. As work progresses there are more and more blocks occurring which are *not* associated with GSR's, while the fewer occurring GSR's are becoming increasingly associated with blocks. This apparently contradictory relation becomes clearer if we assume that there are at least two types of blocks: galvanic and nongalvanic, i.e., one type of block accompanied by greater parasympathetic activity and another type not resulting in or accompanied by increased parasympathetic activation. At the same time, the GSR's are becoming decreasingly associated with spurts in work rate, i.e., momentary speeding up in color naming. Thus as the hour of mental work continues, the GSR is occurring more frequently with inefficient responding and less frequently with efficient responding. There seems to be a shift in the pattern of color-naming responses in its demands on the parasympathetic system.

To discern more specific causality than mere coincidence between the block and the GSR, an examination was made to note the percentage of GSR's which are preceded by blocks within one to five seconds of the galvanic change. Specific causality is assumed on the basis of temporal relation. This examination revealed that 48 per cent of the GSR's are "caused" by blocks and that there is an increasing incidence of such "caused" GSR's as work progresses. This percentage was sufficiently high to indicate a causal relationship, for the GSR may be elicited by numerous other stimuli in the total mental work situation. Momentary increases in work rate, any considerable change in attention, atypical respirations, internal stimuli of various sorts—"mental" or physiological—extraneous laboratory stimuli may, and do, elicit GSR's. In three of the subjects, atypical respirations were responsible for a good percentage of GSR's. The evidence appears to be sufficient to say that most GSR's are caused by stimuli inherent in the mental work, primarily the block.

A question may be asked as to the nature of the blocks that are associated with GSR's and those that are not. Primarily the large blocks and "block groups" are associated with GSR's, while the occasional block does not result in a change in the conductance level. The GSR accompaniment of large blocks and block groups is the result of disturbance to the subject at a perceptual level. Subjects report they are vaguely aware of a "color" in the window of the psychergometer, and later recall it was red, but at the moment are unable to make a correct discrimination to press the correct key. They are forced to reorientate their attention to the color and the keyboard. Attention is shifted effortfully to the task by the work continuity set. The GSR is known to be elicited regularly by shifts in attention (10). If the shift is prolonged, or attitudinal, as a spurt in motivation, the general level of skin conductance is elevated to a new level of excitation background and work rate tends to be increased. It will be remembered that the general trend during mental work is toward a reduction in skin conductance, accompanied by a reduction in work rate. Thus as work progresses the blocks become relatively less and less the result of shifts in attention.

What is the nature of those blocks *not* associated with the GSR? The lesser or occasional blocks tend to be free of GSR association. Although they occur quite frequently in the early period of mental work they increase in frequency as work progresses. They seem to be "automatic" in their nature, i.e., the subject is unaware of their occurrence. The subject responds and continues with the task without any alteration in his effort, as though the block were a normal variation in color-naming behavior. This faltering in response may be due to faulty and inefficient operation of the implicit recognition process resulting from diminishing attention, and the result of competition among the response neural pathways. The general level of skin conductance diminishes and the incidence of GSR's decreases. The higher level of attention is accompanied by a higher level of conductance. The greater tension during the early periods of work may be a factor aiding the awareness and attention. As tension and attention diminish, the "automatic" block becomes more frequent. The diminishing tension (decreasing conductance) throughout the color-naming work presumably removes the reinforcing support needed to maintain the directive set (attention, etc.), thus resulting in a loss of control and more blocks.

IV. SUMMARY AND CONCLUSIONS

A study of some of the physiological variables of mental work was conducted to obtain answers to three problems. (*a*) Do some internal physiological responses change significantly under conditions of light homogeneous mental work, and, if so, what is the nature of the changes? (*b*) Do the changes in some physiological responses show any degree of concomitance with the decrement in mental work production? (*c*) Are there any patterns of internal physiological responses related to mental work products?

The mental work task was color naming. Five colors were presented serially in random order by means of a Bills-type psychergometer (6). The colors were responded to manually at a rate determined by the practiced facility of each subject. Continuous records were made of the changes in rate of respiration and heart beat, relative blood pressure, finger volume, palmar skin conductance, and galvanic skin responses. These physiological and color-naming responses were photographically recorded by means of a Darrow-type photopolygraph (9).

Ten practiced, laboratory-adjusted, paid upper classmen and graduate students were used as subjects. For each subject there were designed four days of color naming by means of manual response, Days 1, 3, 6, and 10; two days of random tapping of the five color-naming keys, while blindfolded, at a rate judged by each subject to be his usual color-naming speed, Days 2 and 4; and two days of passively looking at the five random and serially presented colors at the rate of 90 per minute, Days 5 and 7. Each work period for all experimental conditions was 55 minutes long, preceded and followed by a ten-minute rest period. Each subject was exposed to two days of 55 minutes each of control rest consisting of a state of mental repose and relaxation. All the mental work days were at least two calendar days apart; in fact, most of the experimental sessions were at least two calendar days apart.

The three experimental conditions were designed to satisfy the historical demands of the "central theorists" (i.e., those who would segregate central nervous system activity from supporting metabolic or muscular activity). The keyboard-operating condition was presumed to elicit the motor component (MC), and the color-sensing condition the sensory component (SC) of the mental work condition (MW) of color naming by manual response. To satisfy the central theorists the MC and SC physiological effects were subtracted from the MW condition effects to obtain the residual effects due to the so-called central component. However, all the physiological effects of work of each experimental condition were compared to the effects of the control rest condition and the pre- and postrest periods.

1. The trends of changes in the rate of color naming and blocking are not due to chance but are due to factors present in the mental work operation and the organization of the mental worker.

2. The average rates of respiration and heart beat during color-naming work are significantly greater than the average rates during a control period of mental

repose or a period of rest before and after the work.

3. The average rates of respiration and heart beat are significantly greater during the mental work of color naming than during either the continuous work of random pressing of a psychergometer keyboard or gazing at serially presented colors.

4. Changes in rates of respiration and heart beat during the mental work probably represent momentary adjustments made to the demands of energy mobilization in response to the changing work situation.

5. A significant residue of a half of a respiration and about four heart beats per minute remains after subtracting those effects due to the so-called motor and sensory components of the mental work of color naming as represented in mere keyboard operating and mere "sensing" of serially presented colors.

6. Relative blood pressure significantly increases during color-naming work but significantly decreases during continuous keyboard-operating work and during continuous passive sensing of serially presented colors.

7. The level of palmar skin conductance is significantly greater during color-naming work than during a period of repose, or during a period of rest before and after the work, and greater than during continuous keyboard operating or continuous color sensing of serially presented colors.

8. The level of palmar conductance maintained during the different forms of work is related to the task difficulty, the degree of alertness, the amount of bodily tension, and the amount of energy mobilized to respond to the task situation.

9. A true control base line for physiological activities of minimal mental alertness is difficult to maintain for young adults for a period longer than ten minutes without the intrusion of sleep. All records of physiological responses of an hour of mental repose show a continuous slight decline for the first half hour before the true base line of physiological activity appears.

10. The changing rates of respiration and heart beat, and the changing level of palmar skin conductance during continuous color-naming work are positively correlated with the changing rate of color naming and negatively correlated with the changing frequency of blocking.

11. Changing physiological activities reflect changes in energy mobilization. As color-naming work output decreases, the amount of energy mobilized decreases. Mental work output is a function of the momentary mobilization of energy, the momentary and previous conditions of the circulatory and neuromuscular systems, and the momentary receptivity of the subject to further stimulation.

12. The changing association between galvanic skin responses and color-naming blocks indicates the existence of at least two types of blocks: galvanic and nongalvanic.

13. Galvanic blocks are large blocks or "block groups" which result from momentarily shifted attention and reduced alertness. The subject is quite aware of this block and is forced by the task pressure to reorient attention to the immediate stimulus.

14. The nongalvanic block is more automatic and spontaneous in its production. The subject responds as though the blocked response were a normal variation in color-naming reaction time. This type of block is more a function of response competition than is the galvanic block. As alertness diminishes, loss of control is evidenced by the increasing frequency of nongalvanic blocks.

REFERENCES

1. BARTLEY, S. H., & CHUTE, ELOISE. *Fatigue and impairment in man.* New York: McGraw-Hill, 1947.
2. BENEDICT, F. G., & BENEDICT, CORNELIA GOLAY. Mental effort in relation to gaseous exchange, heart rate, and mechanics of respiration. Washington: Carnegie Institution of Washington, 1933. *Carnegie Inst. Wash. Pub.*, No. 446.
3. BILLS, A. G. Blocking: a new principle of mental fatigue. *Amer. J. Psychol.*, 1931, **43**, 230-245.
4. BILLS, A. G. Causal factors in mental blocking. *J. exp. Psychol.*, 1935, **18**, 170-185.
5. BILLS, A. G. Facilitation and inhibition in mental work. *Psychol. Bull.*, 1937, **34**, 286-309.
6. BILLS, A. G. A new psychergometer. *J. gen. Psychol.*, 1936, **14**, 487-489.
7. COURTS, F. A. Relations between muscular tension and performance. *Psychol. Bull.*, 1942, **39**, 347-367.
8. DARLING, R., & DARROW, C. W. Determining activity of the autonomic nervous system from measurements of autonomic change. *J. Psychol.*, 1938, **5**, 85-89.
9. DARROW, C. W. The Behavior Research photopolygraph. *J. gen. Psychol.*, 1932, **7**, 215-219.
10. DARROW, C. W. The functional significance of the galvanic skin reflex and perspiration on the backs and palms of the hands. *Psychol. Bull.*, 1933, **30**, 697-698.
11. DARROW, C. W. Palmar skin conductance (sweating), an accessory to the activity of the postural mechanism. *Psychol. Bull.*, 1935, **32**, 720-721.
12. DARROW, C. W., & FREEMAN, G. L. Palmar skin-resistance changes contrasted with non-palmar changes, and rate of insensible weight loss. *J. exp. Psychol.*, 1934, **17**, 739-748.
13. DARROW, C. W., & HEATH, L. L. Reaction tendencies relating to personality. In K. S. Lashley (Ed.), *Studies in the dynamics of behavior.* Chicago: Univer. of Chicago Press, 1932. Pp. 59-261.
14. DAY, M. E. The influence of mental activities on vascular processes. *J. comp. Psychol.*, 1924, **3**, 333-377.
15. FREEMAN, G. L. The spread of neuro-muscular activity during mental work. *J. gen. Psychol.*, 1931, **5**, 479-494.
16. FREEMAN, G. L., & DARROW, C. W., Insensible perspiration and the galvanic skin reflex. *Amer. J. Psychol.*, 1935, **3**, 55-63.
17. FREEMAN, G. L., & GIFFIN, L. L. The measurement of general reactivity under basal conditions. *J. gen. Psychol.*, 1939, **21**, 63-72.
18. FREEMAN, G. L., & SIMPSON, R. M. The effect of experimentally induced muscular tension upon palmar resistance. *J. gen. Psychol.*, 1938, **18**, 319-325.
19. GELDREICH, E. W. *Some physiological concomitants of mental work.* Unpublished Ph.D. thesis, Univer. of Chicago, 1952.
20. GELLHORN, E., DARROW, C. W., & YESINICK, L. Effect of variation of blood pressure on the autonomic nervous system. *Proc. Soc. exp. Biol., N.Y.*, 1940, **43**, 236-240.
21. GILLESPIE, R. C. The relative influence of mental and muscular work on the pulse-rate and blood pressure. *J. Physiol.*, 1924, **58**, 425-432.
22. GOLLA, F. L., & ANTONOVITCH, S. The respiratory rhythm in its relation to the mechanics of thought. *Brain*, 1929, **52**, 491-509.
23. KIRBY, T. K. *Curve of recovery from fatigue.* Unpublished Ph.D. thesis, Univer. of Cincinnati, 1942.
24. ROBINSON, E. W. Work of the integrated organism. In C. Murchison (Ed.), *Handbook of general experimental psychology.* Worcester, Mass.: Clark Univer. Press, 1934. Pp. 571-650.
25. ROUNDS, G. H., SCHUBERT, H. J. P., & POFFENBERGER, A. T. Effects of practice upon the metabolic cost of mental work. *J. gen. Psychol.*, 1932, **6**, 65-79.
26. ROWLES, E., & PATRICK, J. R. The effect of various stimuli on the basal metabolic rate, the blood pressure, and the galvanic reflex in man. *J. exp. Psychol.*, 1934, **17**, 847-861.
27. SCHUBERT, H. J. P. Energy cost measurements on the curve of work. *Arch. Psychol.*, 1932, **139**, 1-62.
28. SEARS, R. Psychogalvanic responses in arithmetical work. *Arch. Psychol.*, 1933, **155**, 1-62.
29. SKAGGS, E. B. Changes in pulse, breathing, and steadiness under conditions of startledness and excited expectancy. *J. comp. Psychol.*, 1926, **6**, 303-315.
30. WENGER, M. A. Some relationships between muscular processes and personality and their factorial analysis. *Child Develpm.*, 1938, **8**, 261-276.

(Accepted for publication February 9, 1953.)

Vol. 67, No. 9 Whole No. 359, 1953

Psychological Monographs: General and Applied

A Model of the Auditory Threshold and Its Application to the Problem of the Multiple Observer[1]

Moncrieff Smith
University of Washington

and Edna A. Wilson[†]
Massachusetts Institute of Technology

I. THE THEORY OF THE MULTIPLE OBSERVER

It has often been suggested that two or more independent observers would form a more sensitive detection unit than any one of them alone. This prediction is based on the fact that the sensitivity of an individual varies from moment to moment, so that if one observer is momentarily insensitive, another may detect the signal. In the simplest case, if each of two independent observers has a probability of $\frac{1}{2}$ of hearing a signal, then the probability that both will *fail* to hear it is $\frac{1}{2} \times \frac{1}{2} = \frac{1}{4}$, and the probability that one, the other, or both will hear it is $\frac{3}{4}$.

Schafer (6) has presented the beginnings of a theory of the multiple observer for the case of the masked auditory threshold and has presented data to indicate that in the actual situation multiple observers do not quite come up to predicted improvement. There are a number of reasons for this, the chief one being (as Schafer pointed out) that observers are rarely independent in their judgments, even if they are physically separated from one another. For one thing, if there is a noise background, they listen to the same noise. For another, an experimental test usually involves a long series of signals which are the same for all observers. A number of investigations have shown that human subjects have predictable biases in generating a series of responses. One study (7), in particular, has shown that a medium-strength signal is less likely to be reported if it follows a strong signal than if it follows a weak signal. This tendency also produces correlation among the observers (in an experimental test, at least). Also group morale, fatigue, etc. play a role.

It is obvious, therefore, that a theory of the multiple observer must consider the consequences of correlation among observers. It is easy to see that if the correlation is perfect, and all observers make exactly the same report at the same time, any one of them will be exactly as good as the group. The maximum gain comes when each observer's reports are independent of the reports of all the others (for a constant-strength signal, if negative correlations are not admitted). The theory must also examine the effects of individual differences among the observers with respect to their threshold and its standard deviation. Again it is obvious that if one observer is much more sensitive than the rest, he will do most of the reporting for the group, and the rest of the group could be discarded.

Another aspect of the problem—one that has not previously been considered—is of considerable importance. Human observers are more sensitive detectors if they are allowed to make false reports. That is, their threshold goes down, even when it is corrected for the amount of guessing implied by the false report rate. This gain is of little practical significance, however, unless some means can be found for bringing the false report rate back to an allowable level. One possible way of doing this is to allow each of a group of observers to make false reports, but to demand a high degree of agreement among members of the group before taking action.

---

[1] Research conducted at Lincoln Laboratory, Massachusetts Institute of Technology. The research in this document was supported jointly by the Army, Navy, and Air Force under contract with the Massachusetts Institute of Technology.

[†] Many people have contributed to the conducting of this experiment and to the development of the theory. There are a few we would like particularly to mention.

Dr. Bert F. Green has given us several very helpful suggestions, and we are particularly indebted for his technique of integrating the multivariate normal correlation surface.

Dr. J. C. R. Licklider supervised the calibration of the equipment.

The following people contributed to the design or construction of the equipment: Roy Sallen, Rudy Schreitmueller, Bob Silva, Josiah Macy, and Jack Flannery. We are very grateful for their help.

The theory, therefore, should take account of (*a*) the variation between observers (as well as the variability of each of them around his own threshold), (*b*) the correlation between observers, and (*c*) the implications and consequences of false reports. The case in which there are no false reports by individual observers is simpler, and it will be examined first.

A. INDEPENDENT OBSERVERS—NO FALSE REPORTS

1. *The Case of Two Equally Sensitive, Equally Variable, Uncorrelated Observers*

The usual model of a threshold assumes that some characteristic of the observer varies in time, rapidly enough that its positions on two successive trials are independent. This sensitivity is assumed to vary normally about a constant mean, usually with respect to the logarithm of signal strength. The psychophysical method of constant stimuli is assumed to display the integral of this distribution.

This model of the threshold works only when there are no false reports. A signal of zero strength is infinitely far down on a logarithmic scale, and if such a signal yields a finite probability of report, the area under the threshold distribution cannot be finite. However, if observers try very carefully to avoid false reports, experimental data gathered under many different conditions approximate this model fairly well, and we will use it for an initial examination of the multiple-observer situation.

For the assumptions made here the thresholds of the two observers are represented by two identical normal distributions. Their joint distribution can then be assumed to be represented by a normal correlation surface, with $r = 0$. To find the probability that one or the other of the observers will report a signal of strength S, we want the probability that one or both of the threshold values will be less than S, and this is one minus the double integral of the joint distribution from S to ∞ on each variable. This integral is a function only of its lower limit, hence as S varies from weak to strong, $f_c(S)$ plots the cumulative distribution. Since the means and σ's of the two distributions are assumed to be the same, S may be expressed as a deviation from the mean, divided by σ, and the cumulative distribution function of S is:

$$(1)\quad f_c(S) = 1 - \int_S^\infty \int_S^\infty \frac{1}{2\pi} e^{-\frac{x^2+y^2}{2}} \, dx dy = 1 - \int_S^\infty \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx \int_S^\infty \frac{1}{\sqrt{2\pi}} e^{-\frac{y^2}{2}} dy = 1 - \left[\int_S^\infty \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx \right]^2 .$$

This last expression is the square of the integral of the univariate normal distribution function, and can be obtained easily from tables.

The probability density function is given by the derivative of $f_c(S)$ with respect to S. It is:

$$(2)\quad f(S) = 2 \frac{1}{\sqrt{2\pi}} e^{-\frac{S^2}{2}} \int_S^\infty \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx .$$

It should be noted that this distribution is the same as the distribution of the lower of two values chosen independently from the same normal distribution. The distribution is approximately symmetrical with a mean of $-.564$ and σ of $.826$.

As the number of independent, equal observers increases, the distribution of the lowest of the set of observers shifts down farther from the parent population. The cumulative distribution function will be the same form, but the exponent on the brackets will be equal to the number of observers. The means and σ's of these distributions have been tabulated by Hastings, Mosteller, Tukey, and Winsor (1) not only for the lowest, but for all orders in groups up to size 10. The mean for the lowest of three observers is $-.846$; for five, -1.163; and for 10, -1.539. The σ's for the distributions of the lowest decrease to $.587$ for $N = 10$.

2. *The Case of* N *Independent Observers: No Restrictions on Individual Means and Standard Deviations*

The formulas for the cumulative distribution of the best of N observers follow readily from the argument given above. The only difference is that we can no longer make the same transformation of S for all observers, since each is allowed to have a different mean and σ. If $M_1, \ldots, M_n$ represent the means of the observers and $\sigma_1, \ldots, \sigma_n$ the corresponding σ's, then the probability that at least one of N observers will respond to a signal of strength S is given by:

$$
\begin{aligned}
(3)\quad f_c(S) &= 1 - \int_S^\infty \int_S^\infty \cdots \int_S^\infty \left(\frac{1}{2\pi}\right)^{N/2} e^{-\frac{1}{2}\left[\left(\frac{X_1-M_1}{\sigma_1}\right)^2 + \cdots + \left(\frac{X_n-M_n}{\sigma_n}\right)^2\right]} dx_1 \cdots dx_n \\
&= 1 - \int_S^\infty \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{X_1-M_1}{\sigma_1}\right)^2} dx_1 \cdots \int_S^\infty \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{X_n-M_n}{\sigma_n}\right)^2} dx_n \,.
\end{aligned}
$$

If we perform on each variable X_i, the transformation $z_i = (X_i - M_i)/\sigma_i$ we have:

$$
(4)\quad f_c(S) = 1 - \int_{\frac{S-M_1}{\sigma_1}}^\infty \frac{1}{\sqrt{2\pi}} e^{-\frac{z_1^2}{2}} dz_1 \cdots \int_{\frac{S-M_n}{\sigma_n}}^\infty \frac{1}{\sqrt{2\pi}} e^{-\frac{z_n^2}{2}} dz_n \,.
$$

This distribution can be plotted, with a little effort, from tables of the integral of the normal curve, for given values of the means and σ's.

In order to discuss the question of the set of means and σ's that maximizes the difference between the sensitivity of the group and the sensitivity of the best observer in the group, it is necessary to consider the concept of "best observer." The assumptions we have made so far allow two observers to have the same mean but different σ's, while both are normally distributed. If this condition were realizable, we would have the situation illustrated in Fig. 1. Observer A would be more sensitive to signals above the common mean, but observer B would be more sensitive to weak signals, and the combination would be very effective.

However, common sense indicates that this situation should be very difficult to obtain and experimental evidence confirms the opinion. In realizable situations, the σ's of individual subjects are not very different, and any considerable

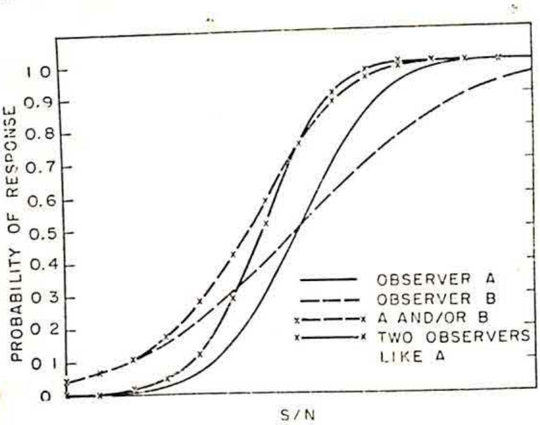


FIG. 1. Report probabilities for two observers with the same mean but different σ's. Curves marked with crosses show report probabilities for teams of two observers.

increase in σ is usually accompanied by a considerable departure from normality of distribution. As a consequence, the assumption of unrestricted variation of σ's is unreasonable, and can be abandoned without much loss of generality of application.

If we restrict our assumptions so that only the means are variable, and if we fix the lowest mean, then it can be seen from formula 4 that the probability of a report by at least one observer will be maximized if all have this mean. This may be demonstrated as follows. Suppose M_1 is the lowest mean. Then the difference between observer one (the best observer) and the group, for any value of S, is given by:

$$(5)\qquad \int_{\frac{S-M_1}{\sigma_1}}^{\infty} \frac{1}{\sqrt{2\pi}}\, e^{-\frac{z_1^2}{2}}\, dz \left[1-\int_{\frac{S-M_2}{\sigma_2}}^{\infty} \frac{1}{\sqrt{2\pi}}\, e^{-\frac{z_2^2}{2}}\, dz \cdots \int_{\frac{S-M_n}{\sigma_n}}^{\infty} \frac{1}{\sqrt{2\pi}}\, e^{-\frac{z_n^2}{2}}\, dz\right].$$

Now $M_2, \ldots, M_n$ are chosen to maximize 5, with the restrictions that $M_1 \leq M_2, \ldots, M_n$ (and all σ's equal). Since each of the integrals must have a value less than one, regardless of the value of S or of the M's, the expression in the brackets will always be positive. Consequently, the product of the integrals in the brackets must be minimized. Whatever the value of S, as M_i decreases, $(S - M_i)$ increases, and the value of the integral decreases. But since $M_2, \ldots, M_n$ cannot be smaller than M_1, they must all equal M_1 to maximize 5.

B. CORRELATED OBSERVERS—NO FALSE REPORTS

In the introductory paragraphs it was indicated that the major sources of correlation among observers are in the situation (the external noise, etc.). However, if we retain the assumption of a threshold that varies up and down in a normally distributed fashion, we must treat the situation as if the thresholds of the subjects actually covaried. We will later substitute what we believe to be a more adequate set of assumptions, but the assumption of covariation of thresholds is satisfactory for an introduction to the role of correlation. If observers make no false reports the picture is adequate. If variation in noise is the major source of correlation, then it makes no difference whether we consider the signal-to-noise ratio to be varying uniformly for all observers, or consider the individual thresholds to have a tendency to vary together.

The general method of deriving the cumulative distribution of the group of observers is the same as before. The probability that at least one observer will report a given signal is given by one minus the probability that none will report, or one minus the multiple integral of the normal correlational surface. However, when the correlations between variables are not zero, the expression for the normal correlation surface is formidable.

For economy of expression, let:

$\sigma_{ij} = \sigma_i^2$ when $i = j$ (this is the variance of the i^{th} observer)

$\sigma_{ij} = r_{ij}\sigma_i\sigma_j$ when $i \neq j$ (r_{ij} is the correlation between the i^{th} and j^{th} observer).

Letting i and j take all possible values from 1 to N, we have the variance-covariance matrix (a_{ij}). The inverse of this matrix is denoted by (a^{ij}) and the determinant of the inverse by $|a^{ij}|$. The multivariate normal correlation surface is:

$$f(x_1, \cdots, x_n) = \left(\frac{1}{2\pi}\right)^{N/2} \sqrt{|a^{ij}|}\; e^{-\frac{1}{2}\sum_{1}^{N}\sum_{1}^{N} a^{ij}x_ix_j} .$$

The a^{ij} in the exponent denotes the elements of the inverse of the variance-covariance matrix. The x's are expressed as deviations from their own means. A more complete treatment is given in Mood (2).

This expression is cumbersome to handle unless some simplifying assumptions are made. Without loss of generality the means and σ's of all observers may be assumed to be equal, since this represents a linear transformation of each variable, and can be compensated for by a change in the limits of integration. It does constitute a loss of generality to assume all intercorrelations of observers to be the same, but if all observers are assumed to be equal in sensitivity and in variability, the additional assumption of equal r's is not difficult to make.

We are willing to make the assumption of equality of intercorrelations, but even so, the expression is a very difficult one to handle. The usual methods of expanding in a series fail, since the series becomes completely unmanageable after a few terms. For $N = 2$ the problem is not too difficult, and values of the integral have been tabled by Pearson (5). Pearson (4) has also published a very ingenious expansion in powers of r for larger values of N, but the series blows up before it begins to converge. However, a method proposed to us by Dr. Bert F. Green allows a feasible, though tedious, solution. It will work in the general case, but is relatively simple only when we integrate up to the same value on each variable, which in this case means that we assume all observers to have the same mean and standard deviation.

The method consists of converting a correlational surface with n variables into a zero-correlation surface of $n + 1$ variables. For five variables, equally intercorrelated, and integrated up to the same z score on each, the development is as follows:

Take five normally distributed variables, $X_1 \ldots X_5$, each with mean zero and standard deviation of σ_x and all intercorrelations equal r.

Now take six normally distributed, independent variables $Y_1 \ldots Y_6$, all with mean equal to zero. Let the standard deviations of $Y_1 \ldots Y_5 = 1$ and of $Y_6 = \sigma_6$. Let

$$X_1 = Y_1 + Y_6$$
$$X_2 = Y_2 + Y_6$$
$$X_3 = Y_3 + Y_6$$
$$X_4 = Y_4 + Y_6$$
$$X_5 = Y_5 + Y_6$$

The correlation between X_i and X_j is the percentage of common variance, or

$$r = \frac{\sigma_c^2}{\sigma_x^2} = \frac{\sigma_6^2}{1 + \sigma_6^2}\ ; \text{ and } \sigma_6^2 = \frac{r}{1 - r} .$$

Now the integral of the normal correlational surface from $-\infty$ to $x_i = a$ for all i's is the probability that the following conditions will be met simultaneously.

$$\begin{array}{lcl} Y_1+Y_6<a & & a-Y_1>Y_6 \\ Y_2+Y_6<a & & a-Y_2>Y_6 \\ Y_3+Y_6<a & \text{or:} & a-Y_3>Y_6 \\ Y_4+Y_6<a & & a-Y_4>Y_6 \\ Y_5+Y_6<a & & a-Y_5>Y_6 \end{array}$$

Here we have six normally distributed, independent variables, and ask the probability that five of them are greater than a specified sixth. ($u_1 = a - Y_1$, $u_2 = a - Y_2, \ldots, u_5 = a - Y_5$) have a mean of a and σ of 1, but Y_6 has mean equal to zero and $\sigma = \sigma_6$. The desired probability is the probability density of Y_6 taking some value t, and all the others ($u_1, \ldots, u_5$) exceeding t, integrated over all values of t, or:

$$P = \int_{t=-\infty}^{\infty} \left[\frac{1}{\sigma_6\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{t}{\sigma_6}\right)^2} \left\{ \int_t^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(u-a)^2} du \right\}^5 \right] dt$$

$$= \int_{-\infty}^{\infty} \left[\frac{1}{\sigma_6\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{t}{\sigma_6}\right)^2} \left\{ \int_{t-a}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} dz \right\}^5 \right] dt \cdot$$

This expression can be integrated numerically to give the probability that all the X's (correlated values) will fall below some specified value a. The process is tedious, but taking t in .1 steps and summing products seems to provide fair accuracy to the third decimal place.

C. THE INTERPRETATION OF FALSE REPORTS

All of the foregoing discussion is based on the implicit assumption that the observer either does or does not have a conscious sensation of tone, and that he reports only when he has it. His sensitivity is assumed to vary in time, so that the objective signal strength that leads to the experience is not always the same, but his probability of hearing extremely weak signals is infinitesimally small.

Within this framework it is difficult to account for the instances in which the observer reports that he hears a signal that was not physically present. It could be assumed that forcing a more liberal reporting attitude on the subject merely causes him to report positively in some fixed proportion of the trials on which he hears no tone. Consequently, a correction for guessing, based on the percentage of false reports, should yield the same threshold distribution that is given when there are no false reports. Preliminary results, however, showed that this was not the case.

It seems desirable to find a new framework in terms of which the threshold can be discussed. Referring to the conscious experience of tone adds little to the analysis of the threshold, and it seems better to view the report of the observer simply as a bit of behavior, determined jointly by the instructions, the signal, and his own internal variability. The details of this analysis are so closely linked to the data to be presented that further discussion at this point is unprofitable.

II. THE EXPERIMENTS

The plotting of the relationship between signal strength and relative frequency of report of the signal requires great quantities of data, gathered by the method of constant stimuli. Moreover, to get even a relatively stable result from instructions, particularly with naive observers, it is necessary to keep the observer informed of his results. Both of these requirements are difficult to meet without the aid of automatic machinery.

The main experiment was designed to permit five physically separated observers to listen simultaneously to the same 800-cycle tone (masked by a broad-band noise) and to report independently whether or not they heard it. Twenty-three groups of five men each participated in the experiment. Twelve of these groups were instructed to indicate that they heard the tone only when they were quite certain of it. The other eleven groups were instructed to report whenever they *thought* they heard the tone. Thus, data were gathered for units of five observers under two different attitudes toward responding.

A. EQUIPMENT

1. *General*

Figure 2 shows a block diagram of the apparatus for presenting signals and recording responses. The unique feature of this apparatus is the automatic signal selection. The multiple-tap attenuator had 16 output leads, each carrying the signal, but with a different degree of attenuation. These ranged from zero attenuation through 14-db attenuation in one-db steps ("db," of course, stands for *decibel*). The sixteenth output was tied directly to ground and carried no signal.

The 16 signal strengths were fed into a binary fan (see Fig. 2A) which selected one of the 16. The selection was controlled by the pattern of energization of the four relay coils. This pattern was determined by the pattern of holes punched in a teletype tape. The 32 symbols transmitted by teletype may be coded into a pattern of five holes punched (or not punched) in a line across the tape. Since only 16 signal strengths were used, only four of the five rows of holes were necessary. As the tape advanced through the teletype reader, the reader transmitted the pattern of holes to the four relays. For example, if there were holes in the tape in positions 1, 3, and 4, the corresponding relays were closed, and the signal at 13-db attenuation was selected. A sequence of signal strengths was punched into the tape, and the tape was stepped through the reader by an impulse coordinated with the signal timer.

The responses of the observers were also punched into a teletype tape. A teletype perforator was modified by mounting solenoids under the five keys that corresponded to a single punch in one of the five rows. The observer was given a response box with a spring-return toggle switch, and told to push the switch on the trials on which he heard the tone. Each observer controlled one of the five solenoids, so that a performance could be scored by counting the holes in one row of the response tape. In normal operation the teletype perforator advances the tape as soon as any key is depressed, punching out the code corresponding to that key as it does so. To keep the responses to a given signal lined up, the normal perforator circuit was disconnected and a

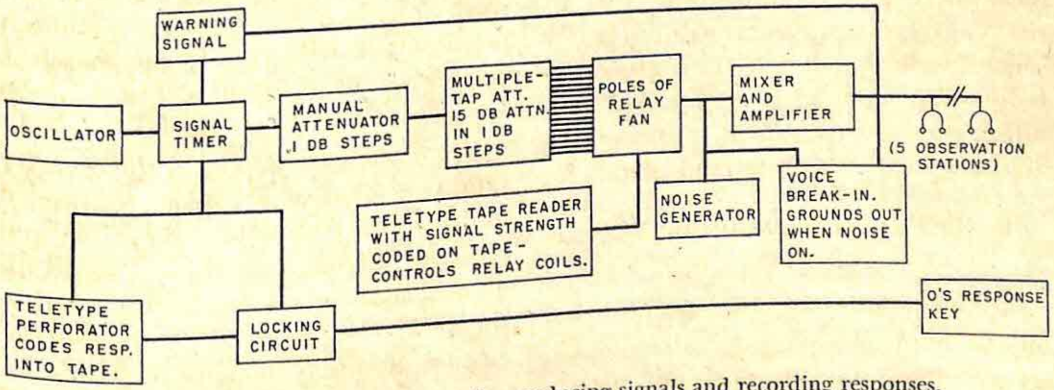

WARNING SIGNAL
OSCILLATOR
SIGNAL TIMER
MANUAL ATTENUATOR .1 DB STEPS
MULTIPLE-TAP ATT. 15 DB ATTN. IN 1 DB STEPS
POLES OF RELAY FAN
MIXER AND AMPLIFIER
(5 OBSERVATION STATIONS)
TELETYPE TAPE READER WITH SIGNAL STRENGTH CODED ON TAPE-CONTROLS RELAY COILS.
NOISE GENERATOR
VOICE BREAK-IN. GROUNDS OUT WHEN NOISE ON.
TELETYPE PERFORATOR CODES RESP. INTO TAPE.
LOCKING CIRCUIT
O'S RESPONSE KEY

FIG. 2. Block diagram of apparatus for producing signals and recording responses.

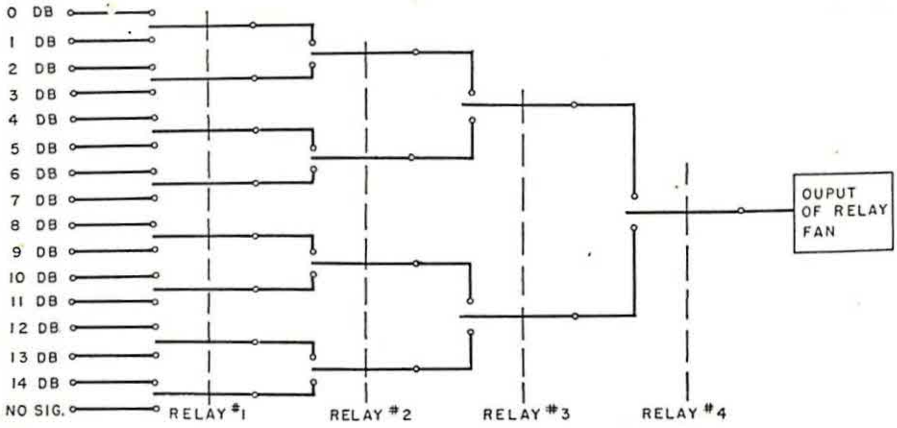


FIG. 2A. Schematic diagram of binary relay fan.

pulse from the signal timer advanced and punched the tape. The observer's response key actually controlled a locking relay, which, in turn, kept the appropriate solenoid energized until the tape was punched. The same impulse that actuated the perforator cleared the locking relays.

The observer's response key was a double-throw switch. For the main experiment, only a yes-no response was required and the two poles were connected together so that the switch could be operated in either direction with the same result. However, in some experiments, four categories of response were required (corresponding to different degrees of certainty that a tone had been presented). For this condition the response-box wiring was changed so that one box controlled two rows of holes on the response tape. The four categories of response were indicated by the observer by throwing the switch both ways, to the right, to the left, or neither. Only two observers could be run simultaneously under these conditions.

Reading the response tapes took additional apparatus. Counting the number of holes in one row of the tape gave the total number of responses, but this total had to be broken down by signal strength to be useful. The information on each response tape had to be broken down into 80 separate counts—the number of times each of five observers responded to each of 16 signal strengths. To do this took a panel of 80 electromagnetic counters (5 columns and 16 rows). In order to record the counts by signal strength, two teletype readers were run synchronously, one reading the signal tape and the other the response tape. The signal-tape reader selected the row of counters to be activated, and a hole in the response tape was recorded on the counter in the corresponding column. The readers were advanced every second, so that a tape containing responses to 180 signals could be scored in three minutes. This device made it feasible to score the performance after each trial and inform each observer of his results before the next trial.

Two other scores were needed. For the evaluation of the group as a joint detection unit, it was necessary to count, by signal strength, the number of times 0, 1, 2, 3, 4, or all 5 observers responded to a given signal. From these counts curves could be plotted showing, for each signal strength, the percentage of times at least one, at least two, etc. observers responded. A special relay circuit, fed by the response tape, made this count possible.

The other count was for the intercorrelation of observers. This, again, had to be made by signal strength, since the correlation is meaningful only if signal strength is partialed out. The count was also made separately for each pair of subjects. Ten different pairs can be drawn from five observers, so each response tape contributed data toward 160 different correlations. The counts that had already been made provided the marginal totals of a two-by-two correlation table for each pair of observers. The only additional information needed was one of the cell frequencies. This was obtained with the help of another relay circuit, which responded to coincidence of punches in a specified pair of rows on the response tape.

2. *Stimulus Data*

a. Timing of signal. For all experiments the signals were separated by a uniform period of five seconds. The duration of the signal was .41 second. The noise was on continuously for the whole period of the test, and each signal (or blank) was preceded by about one second by a warning signal. The warning signal was a sharply damped oscillation that was clearly audible above the noise.

b. The signal. The signal was an 800-cycle pure tone, generated by a General Radio oscillator. Its intensity, for the loudest signal used, was —37.3 db re 1.0 volt. This measurement was made at the earphones, with a Ballantine Voltmeter.

c. The noise. The noise was generated by a 6D4 gas tube, amplified. The spectrum of the noise generator, as measured at the earphones by a Hewlett Packard wave analyzer set for 135 cps half bandwidth, was flat to within one db from 400 to 2500 cps, with a drop of 10 db from the peak down to 100 cps, and a drop of about 12.5 db from the peak up to 10,000 cps.

The noise intensity was measured for a band from 500 to 2400 cps. An amplifier with a high input impedance and a band-pass filter were connected in parallel with the earphones, and the output of the filter was measured with a Ballantine Voltmeter. The rms voltage at the output of the filter was —28.0 db re 1.0 volt. This value represents a correction of the actual Ballantine reading for the fact that the noise peaks were clipped, and a correction for the fact that noise, rather than a sine wave, was being measured. The noise level in a one cps band was, therefore, —60.8 re 1.0 volt, for the 1900 cps band from 500 to 2400 cps.

The signal-to-noise level of the loudest signal was $-37.3 - (-60.8) = 23.5$ db.

d. The earphones. The earphones used were PDR-8's. Their frequency response curves were obtained for a 1 volt constant input with the earphone feeding into a 6-CC coupler. The resulting sound pressure readings were corrected (3) for the earphone cushions that the observers wore.

The sound pressure level was fairly uniform over the range from 100 to 7-8000 cps. Of the ten individual phones used, nine varied less than 4 db in the range from 500 to 2500 cps. The one exception had a sharp peak at 2000 cps, followed by a dip at 2500 cps. This range was 15.5 db, but because of the frequencies at which the deviations occurred, the noise level as calculated above should apply even in this case. None of the phones showed any peculiarities in the immediate region of 800 cps.

At 800 cps the sound pressure levels varied from 100.0 to 107.0 db re one microbar for a 1 volt input. This maximum deviation was between the right and left phones used by observer No. 2. The other pairs were better matched, differing by 1.5, 1.0, 4.0, and 0.5 db.

The loudest signal used was —37.3 db re 1 volt. Hence the sound pressure level for this signal was between 62.7 and 69.7 db re one microbar, depending on the phone used. The noise, in terms of energy per cycle, was 23.5 db below this.

e. The range of signals. The automatic attenuation was supposed to cut the signal intensity by 1-db steps, over a range of 15 db. It failed slightly in doing this. No single step was off by more than 0.1 db, and the maximum cumulative error was 0.3 db. The actual intensities, expressed as decibels attenuation from the loudest signal, were 0, 1, 2, 3, 4, 5, 6.1, 7.1, 8.2, 9.3, 10.2, 11.2, 12.2, 13.1, and 14.2. All graphs expressing results are corrected for these errors.

B. Observers

The observers were all enlisted military personnel from Fort Devens, Massachusetts. They were brought in each morning and returned in the evening, and were available for testing for five or six hours.

Twenty-three groups of five men each participated in the main experiment. Of these 115 observers, 103 were different individuals. In seven instances, men were erroneously returned for a second test period. In addition, one group of five men was intentionally returned for a number of days for another purpose. On their first day they ran under the standard conservative instructions. On the second day they ran under standard liberal instructions. In no case was there evidence of a practice effect from the first day's testing, so these 12 repeat observers were treated in the same way as the rest of the men.

On the whole, the men were remarkably cooperative. Conditions in the listening rooms were far from ideal, particularly for the earlier groups. Some of these were run in unventilated cubicles on a hot summer day. Often the men reported that they were short on sleep, but there were only about a half dozen known instances of observers falling asleep in the listening room. The experimenter made a practice of checking the response tape to correct this situation if it arose. Data from a 24th group were discarded because the members could not be kept awake, but in the other cases, no correction of the results was possible. This factor thus represents a slight source of error.

The Army General Classification Test scores of the observers were obtained from Fort Devens when possible. The average of the known scores was 109—slightly above the mean of the military population.

C. Procedure

When the group of men arrived in the morning, they were given a brief explanation of the purpose of the experiment, followed by detailed instructions on their task. They were then given a demonstration of the sound of the noise and the signal in the noise. At the end of the demonstration all men were questioned to insure that they had recognized the signal, and knew what to listen for. They were then given a practice series, made up in the same manner as the test series. The practice series was scored and the results explained to the men before the experiment proper began. There is evidence that two men still failed to understand (in the early part of the experiment at least) what was expected of them. Since the interest was in the performance of the whole group, the data from these men are included in the totals. Their performance curves for the whole day were not very atypical.

The experiment itself was made up of listening periods of about 15 minutes each. Each listening period was made up of 180 signals (10 of each of 15 different strength signals, and 30 blanks). It was possible to get in 6 to 9 listening periods in the course of one day. One group had 6 periods, 2 had 7, 15 had 8, and 5 groups had 9 periods. Thus the typical observer responded to 1,440 signals, 80 at each of the signal strengths and 240 blanks.

The stimulus tapes were made up in a random sequence, subject only to the restriction that 10 of each signal and 30 blanks occurred on each tape. Every tape the observers heard had a different sequence, but the same set of tapes was used for all groups. Preliminary experience showed that the first few responses on each tape were apt to be unreliable, so each tape began with four extra signals, including one loud one. Observers responded to these four, but the responses were not scored.

Every response tape was scored, and a general picture of the results was given to the observers before the next tape was run. It is believed that this procedure was, in large part, responsible for the high degree of cooperation we obtained from the observers. Typically, they watched the scoring and it was not unusual for strong competition to develop in the group. The immediate knowledge of results also helped considerably in stabilizing the attitude of the observers toward reporting.

The portion of the instructions designed to induce a conservative or liberal attitude toward the reporting of signals is given below. Additional caution or encouragement was given during the course of the day, when it was indicated by the results.

Conservative groups:

Keep in mind that it is important for you to be sure you hear the tone. In many cases there will be no tone, and you will be making a mistake if you indicate that you hear one. None of the tones is really easy to hear, but don't push the switch unless you are sure *that you heard the tone. If you are in real doubt as to whether or not you heard a tone, assume that there was none.*

Liberal groups:

Keep in mind that all of these tones are hard to hear, and that you will rarely be absolutely sure that you heard something. But if you think you heard something, probably you did—report it. If you are very sure that you didn't hear anything—then don't touch the switch.

In addition to the main experiment, two groups of five men each were tested under special conditions. These men were asked to respond with one of four categories of response, instead of the two categories (yes or no) required in the main experiment. These four categories were verbalized as follows:

1. Certain that there was a tone. This corresponds to the "yes" category used by the conservative groups.
2. Think there was a tone. If a response fell in either category 1 or 2, it was assumed to be comparable to the "yes" category of the liberal groups.
3. Didn't hear it, but guess it was there.
4. No tone.

Categories 3 and 4 were intended to break down the "no" category of the liberal groups.

III. RESULTS

A. CUMULATIVE RESPONSE CURVES

1. *Curves for All Observers*

The quantity of data gathered makes it possible to get a good picture of the report rate as a function of signal-to-noise ratio. Figure 3 gives this plot for all observers. The conservative curve is based on the performance of 60 observers, with a total of 4,900 judgments on each signal strength, and 14,700 judgments on blanks. The liberal curve is based on 55 observers, with 4,350 judgments for signals and 13,050 for blanks. Each judgment was given equal weight, although the number of listening periods for an individual varied from six to nine.

The curves approximate a normal ogive, but the approximation is not very close, even for the conservative curve. The major deviation from the ogive lies in the fact that the false report rate is not

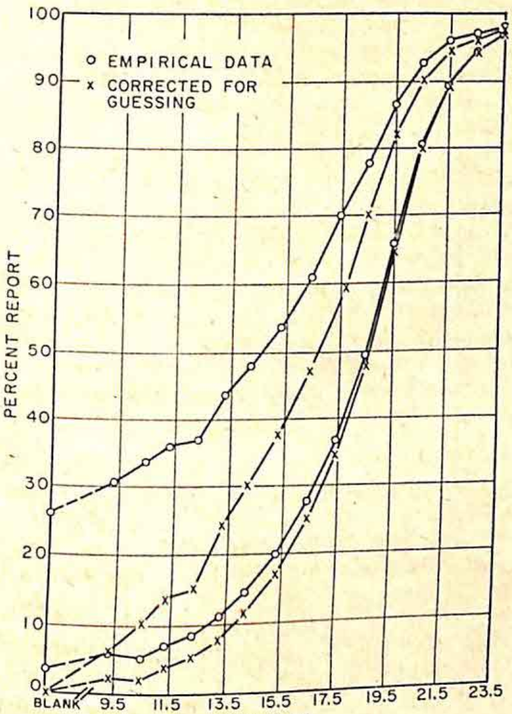


FIG. 3. Averages for all observers in main experiment. Lines marked with crosses are corrected for guessing. The top two lines are the raw and corrected averages for observers given liberal instructions; bottom two lines are for conservative observers.

zero. However, this deviation extends up past the middle range of signals. When the two curves are plotted on normal probability coordinates they still tend toward the shape of an ogive. The upper branch of the curve (above 18.5 db) is an approximation to the normal distribution in the case of the conservative groups, although the two loudest signals were reported somewhat too infrequently for the distribution to resemble the normal very closely. The same is true of the liberal observers, to a greater extent. These curves are known to include some sources of error—partial equipment failures, drowsy or uncooperative observers, etc. However, it was impossible to discard data for individuals in considering the effectiveness of the group, so these curves are presented for comparison with the curves for the group considered as a multiple detection unit.

Each of these curves shows one marked deviation from a smooth line. The liberal curve has a dip at 12.5 db and the conservative curve at 10.5 db. It is difficult to give a meaningful statement of the significance of these deviations, but at least one of them appears to be within the limits of sampling variability. To test this we took individual percentage reports for the questionable point and for the point on either side. Each percentage was subjected to the arc sine transformation, and the variance of individuals was computed for each point. The variances for the three liberal points were tested for homogeneity by Bartlett's test, and so were the three conservative variances. Both sets were found to be homogeneous, so the estimate of the variance of an individual point was based on the variance of the set. The expected percentage in each case was estimated graphically, and converted to an arc sine. The difference between the estimated and the observed values was divided by the appropriate standard error of the mean to yield a t value. For the dip in the conservative curve, this t was .96 and for the liberal curve it was 2.00. For 54 df the 5 per cent level of t is 2.01, so that even the dip in the liberal curve is not quite up to the conventional level of significance. However, these are conservative estimates of significance, since they ignore the correlation between observed and estimated percentages.

There is no reason to suspect the apparatus in the case of either of these dips. The groups were run on alternate days, so any apparatus failure too intermittent to be detected by other means should show up in both groups.

The correction for guessing. If the

liberal observers attain their higher report rates by guessing, it should be possible to detect it from the data. We assume that an observer either hears the signal or he doesn't, but that if he doesn't hear it, he has a certain probability of reporting anyway. If this is the case, by instructing the observers to be liberal in their reporting, we have merely increased the probability of reporting when nothing is heard.

Under these assumptions, when no signal is presented, none can be heard, and the number of false responses gives us the probability of a guess. Let P_t be the true proportion of signals heard, and P_g be the proportion of guesses. Then $(1 - P_t)P_g$ will be the proportion of reports attributable to guesses. If P_o is the observed proportion of responses, then $P_t + (1 - P_t)P_g = P_o$; and $P_t = (P_o - P_g)/(1 - P_g)$. P_t is the corrected proportion of reports.

These corrected percentages are shown in Fig. 3. If the assumptions we made above were true, these corrected curves under liberal and conservative attitudes should be the same. Since they are not the same, the additional reports made under the liberal attitude represent something more than a guess.

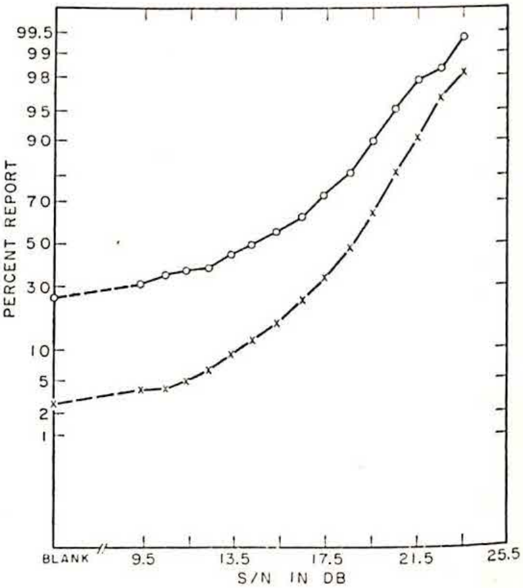


FIG. 4. Average performance of *selected* observers, plotted on normal probability coordinates. Top line is liberal observers, bottom line conservative.

2. *Curves For Selected Observers*

Figure 4 gives a better picture of what the cumulative response distributions should be. It is plotted on normal probability coordinates so that deviations from normality may be seen easily. It was pointed out in the previous section that the average curves for all observers were known to contain some sources of error. An attempt was made to eliminate the more gross sources of error. It was felt that any observer should have been able to report 95 per cent or more of one of the louder signals if he had been awake and alert throughout the experiment. Therefore, we discarded the data of any observer in either group who failed to reach the 95 per cent report level for at least one signal strength. This criterion eliminated 6 from the conservative group and 7 from the liberal group. This criterion may seem too stringent for the conservative group, and the question might be raised as to whether we are discarding observers who lapsed into another state (and hence do not represent the same population as the rest) or whether we are capitalizing on error and discarding the extremes of a continuous distribution. Actually there was a strong tendency in both groups for those who fell below 95 per cent to fall well below it. A few of the observers who were discarded probably represent extreme cases of insensitivity to the masked signals. However, we feel that inattentiveness or actual sleep was responsible for most of these exclusions.

In order to make the selected groups

more homogeneous, an additional cut-off was established for false report rate. Conservative observers with false report rates of 5 per cent or higher were discarded (an additional 13 observers). Liberal observers whose false report rates fell below 15 per cent (two cases) or above 35 per cent (two cases) were also discarded. This left 41 observers in the conservative group, and 44 in the liberal group. Unfortunately, these criteria do not eliminate all sources of error. For example, the dip in the liberal curve at 22.5 db is almost certainly due to occasional equipment failures at this signal strength. Most of this dip comes from one or two groups, and the records showed an extreme tendency for all five observers to miss the same signal. However, it was felt that any further selection of cases on the basis of the records themselves could lead to an unrealistic picture of the process under examination.

The resulting curves (Fig. 4) are not very different from the total averages. The conservative curve is sharpened, that is, it is lower for the weak signals and higher for the strong ones. This, of course, is to be expected from the nature of the selection process. The upper branch of the conservative curve for selected observers more nearly approximates the normal, but still shows a falling off for the strongest signal. The selected liberal observers have the same false report rate as the unselected group, but the rest of the curve is raised. This curve also shows a better approximation to normality in its upper half, but still falls off for the strongest signals.

B. The Performance of Multiple Observers

Cumulative response curves were computed for the group as a detection unit. These curves show, for each signal strength, the percentage of times that at least one observer responded to a particular signal; the percentage of times at least two responded; and so on up to the percentage of times all five of the observers responded. All the reports that go to make up the percentage report for the "two-or-more" curve also go into the curve for "one-or-more." Thus the five curves may have the same ordinate, but they cannot cross one another. These curves are plotted as parts of Figs. 16 and 17, Fig. 17 for the observers working under a liberal attitude toward reporting, and Fig. 16 for observers instructed to report only when they were certain they heard the signal (conservative attitude).

If the reports that go to make up these curves were used in a detection situation, it would be necessary that the false report

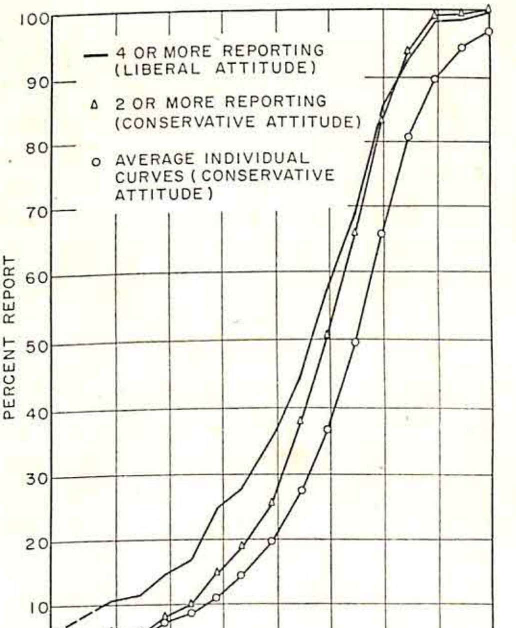


Fig. 5. Comparison of average individual performance with group performance under liberal and under conservative instructions.

rate be low. With observers working under a liberal attitude toward reporting, for example, it would probably be necessary to adopt a criterion of four or more observers reporting. Even this criterion gave a false report rate of 5.7 per cent. The requirement that all five observers report the signal brings the false report rate down almost to zero for the liberal groups, but also has a severe effect on the report of stronger signals.

When false report rates are held constant, there is little difference between the effectiveness of the groups working under the two different attitudes. Fig. 5 shows a comparison of the curves for "four-or-more" reporting under the liberal attitude, and "two-or-more" reporting under the conservative attitude. These curves have approximately the same false report rate, and the cumulative response curves are quite similar. The liberal groups were somewhat more sensitive to weak signals, but this may be due in part to the fact that they had a slightly higher false report rate. The conservative groups were actually slightly superior in detecting the stronger signals.

The curve for "all five" observers reporting under the liberal attitude has a .15 per cent false report rate. To get a lower rate for the conservative observers, we also have to adopt a criterion of all five reporting. In this comparison the liberal group is clearly better. However, the "four-or-more" (.17 per cent) and the "three-or-more" (.78 per cent) criteria yield very low false report rates for the conservative groups. If we compare either of these curves with the comparable liberal curve (all five reporting), we see again that the liberal groups are superior on weak signals, inferior on strong ones.

Figure 5 also compares the two group curves with the average curve for individual observers working under the conservative attitude. The false report rate for the individuals is almost identical to that of the conservative groups (two or more observers reporting) so that a direct comparison is possible. If the effectiveness of the group is measured in terms of the strength of signal which is reported 50 per cent of the time, then the group is not very superior. This gain is only slightly more than one db. However the curves show that this comparison in terms of the conventional threshold does not tell the whole story. The group makes a sharper distinction between signal and no signal than the individual does. If we compare the two on the strength of signal that will be reported 95 per cent of the time, the group is two db better.

C. Intercorrelations among Observers

In the main experiment, five observers listened to, and reported on the same signals. In the introduction to this report it was shown that the effectiveness of the group was dependent, in part, on the tendency for the observers to report, and to miss, the same signals. More than this, the correlations between observers yield an estimate of the proportion of the variance common to all observers. Since the observers were physically isolated from one another, there is no way that one of them could have gained knowledge of the response made by the others to a given signal. As a source of common variance, then, we are left with variation outside the observers —i.e., the masking noise. It is possible that other, and unintended, noises occasionally penetrated the sound-shielded listening booths, the earphone cushions, and the masking noise. These external noises may have made a small contribution to the common variance. Momentary variations in the masking noise,

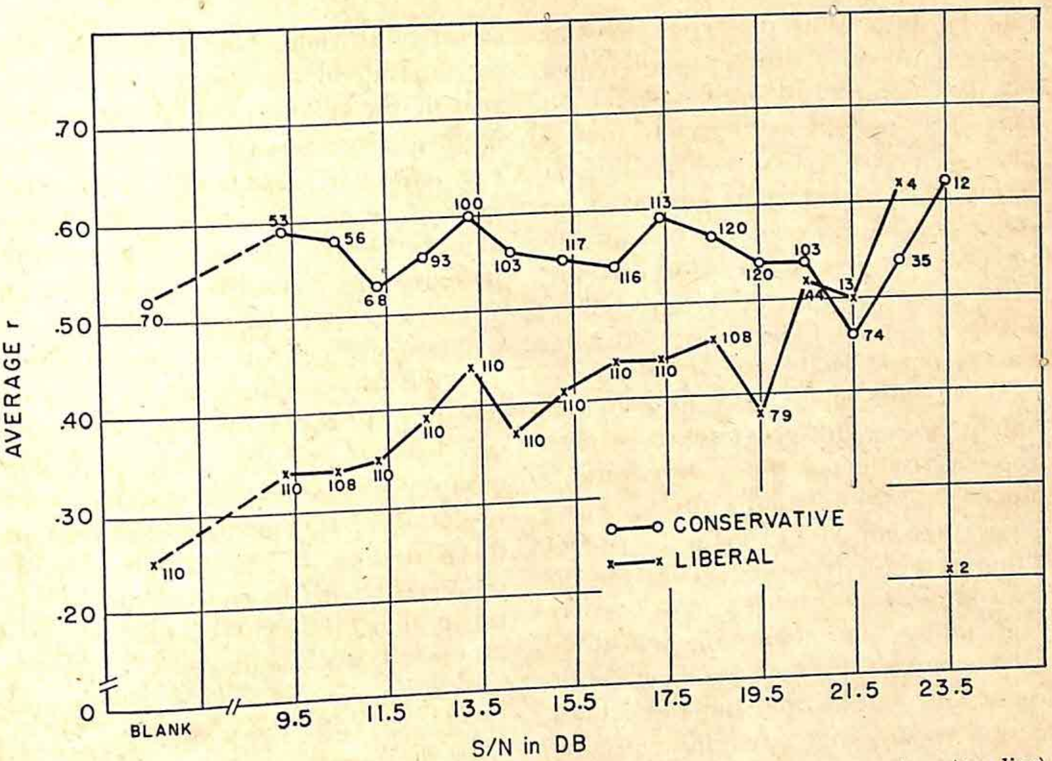


FIG. 6. Average intercorrelations of observers working under conservative instructions (top line) and under liberal instructions (bottom line). Alongside each point is the number of correlations making up the average (120 possible for conservative curve; 110 for liberal).

however, seem to be a more important source of common variance. Other secondary sources include those mentioned in the introduction, fatigue, sequence effects, etc.

The correlations between observers are meaningful only if the effect of signal strength is partialed out, since there is a strong tendency for all observers to report the strong signals and to fail to report the weak signals and the blanks. For this reason, correlations were computed by signal strength.

The data were given in the form of a dichotomy, report or no report, and the tetrachoric correlation coefficient was used. Separate correlation coefficients were computed for each of the ten pairs in each group of five observers, so that each group yielded ten correlations for each of 16 signal strengths. Each individual correlation was based on 60-90 observations for signals, and 180-270 for blanks. The correlations for a given signal strength were averaged by converting each r to Fisher's Z, averaging the Z's, and converting back to r. The resulting average correlations are plotted by signal strength in Fig. 6. Separate plots are given for liberal and conservative observers.

The nature of the data was such that correlations could not always be computed. When the signals were strong, individual subjects often reported every signal. In this case, the correlation is indeterminant. In other instances each of two observers missed one or two signals. If the missed signals were the same for both observers, the correlation was $+1$;

if not, it was −1. Perfect correlations obtained this way can also be considered indeterminant. They were not included in the average.

Thus the correlations for the strongest three or four signals in both groups, and also for the weaker signals in the conservative groups, tend to be quite unstable. Not only are there fewer intercorrelations to average, but each of the individual correlations tends to be unstable because of the extreme marginal split. However, feeling that an unstable estimate was better than none, we computed and averaged every determinable correlation, no matter how extreme the marginal split. The graphs in Fig. 6 show, alongside each point, the number of correlations making up the average.

The correlations between members of conservative groups fell between .5 and .6 for most signal strengths, and the question arises as to whether or not these correlations can be assumed to be drawn from the same population. An accurate answer to this question is extremely difficult to obtain, but we can make a crude estimate. The standard error of the tetrachoric correlation is given as:

$$\sigma_{r_t}=\frac{\sqrt{pqp'q'}}{z_x z_y\sqrt{N}}\sqrt{(1-r^2)\left[1-\left(\frac{\sin^{-1} r}{90^\circ}\right)^2\right]}$$

where p and q represent the marginal proportions of one variable with z_x as the ordinate of the normal curve corresponding to that proportion. The corresponding proportions and ordinate of the other variable are p' and q', and z_y.

The average r is approximately .55. We can make an approximate test of the hypothesis that the various average r's are drawn from a population with $r =$ .55 by the following procedure. We assume that the two marginal distributions are the same, and take the data for these values from Fig. 3. We simplify further by taking a single value of N (the number of cases on which a single tetrachoric r is based). For convenience, we used 81. The standard error of a tetrachoric r of .55 was then computed for the marginal splits at each signal strength. To estimate the standard error of the mean correlation we divided by the square root of the number of correlations on which the average was based. The individual tetrachoric correlations are not normally distributed, but the distribution of the mean of 50 or more of them should tend toward normality, so that the deviations of the means from the over-all mean of .55 were evaluated in terms of the normal distribution.

Three of the 16 average correlations deviated beyond the 5 per cent confidence limits so established. The value of .60 at 13.5 db exceeded the confidence limit of .597. At 17.5 db, r was .59; the confidence limit is .576. At 21.5 db r was .46; the lower confidence limit is .494. If we had real confidence in these limits, we should have to reject the hypothesis that these correlations came from the same population, since, if the probability is .05 that any one will fall outside the limits, the probability (from the binomial expansion) is .047 that 3 or more out of 16 will exceed the limits.

Possible sources of error in this estimate are so great, however, that it can be taken as little more than an estimate of the order of magnitude of the error. The correlations are not all independent as assumed by the test. There is reason to believe that the groups varied among themselves more than might be expected on the basis of the same sort of calculation. If we look at the variability of groups around the average value of .55, the extreme deviations from .55 do not look very great. At 17.5 db, for example, seven groups averaged above .55, one averaged .55, and four were below. At no signal strength were there more than nine of the twelve groups above or below the average r of .55.

It seems reasonable to conclude that we have no good reason for rejecting the hypothesis that the correlation between observers is independent of signal strength when observers adopt the conservative attitude. The conclusion can only be regarded as tentative, but at least there seems to be no evidence in these data for any trend.

Even the evidence for absence of a trend is doubtful. When we exclude from consideration those observers having a false report rate exceeding 5 per cent, and also those who fail to reach 95 per cent report on any of the strong signals, we find some indication of a trend in the figures. For this selected group of conservative observers, the correlations on the weaker signals were higher than those on the stronger signals. In making up this selected group of observers, 19 men were discarded—13 for a high false report rate, and 6 for failing to reach 95 per cent report on any signal. The discards were so arranged in the groups that there were 60 possible intercorrelations among the selected observers. The average intercorrelations, together with the number of individual correlations on which the averages were based, are as follows: .64(31); .72(24); .66(24); .64 (33); .59(46); .62(48); .58(50); .58(57); .56(58); .62(57); .58(60); .54(60); .56(57); .52(38); .60(14); .46(1). These correlations are listed in order of ascending signal strength, with the first one being the correlation when no signal was presented.

It may be seen from these figures that practically all of the correlations were increased with respect to those from the unselected groups. This was to be expected, on the grounds that the selection procedure eliminated those observers who were inattentive, or who failed to follow instructions (too liberal in reporting). The relatively greater increase on the weak signals was not expected, however. It might be argued that our selection on the basis of false report rate eliminated those whose error variance in reporting blanks happened to be high. This would explain the relatively large increase in the correlation for the blanks, but not for the weak signals. There was no explicit selection on the basis of response to these, so that our selection procedure should not have systematically increased these correlations.

For the liberal groups, Fig. 6 shows that the average intercorrelation was not independent of signal strength. There is a definite trend in the correlations, with those for blanks and weak signals being lower than those for the stronger signals. The data for selected liberal observers show the same thing. The selected observers average .07 higher, but this increase is uniform (within reasonable sampling limits) for all signal strengths.

D. Rate of Information

If we consider the individual observer as a device for decoding signals into reports on the presence or absence of a signal, we can arrive at another method of evaluating the effect of the attitude taken by the observer. If the observer adopts a liberal attitude toward reporting, this is analogous to turning up the gain on the last stage of an amplifier system. If the signals being amplified are weak, amplifier noise will be increased along with the signal, and the result may not be helpful. In the case of the observers, more signals are reported, but more blanks are reported as signals, too. The

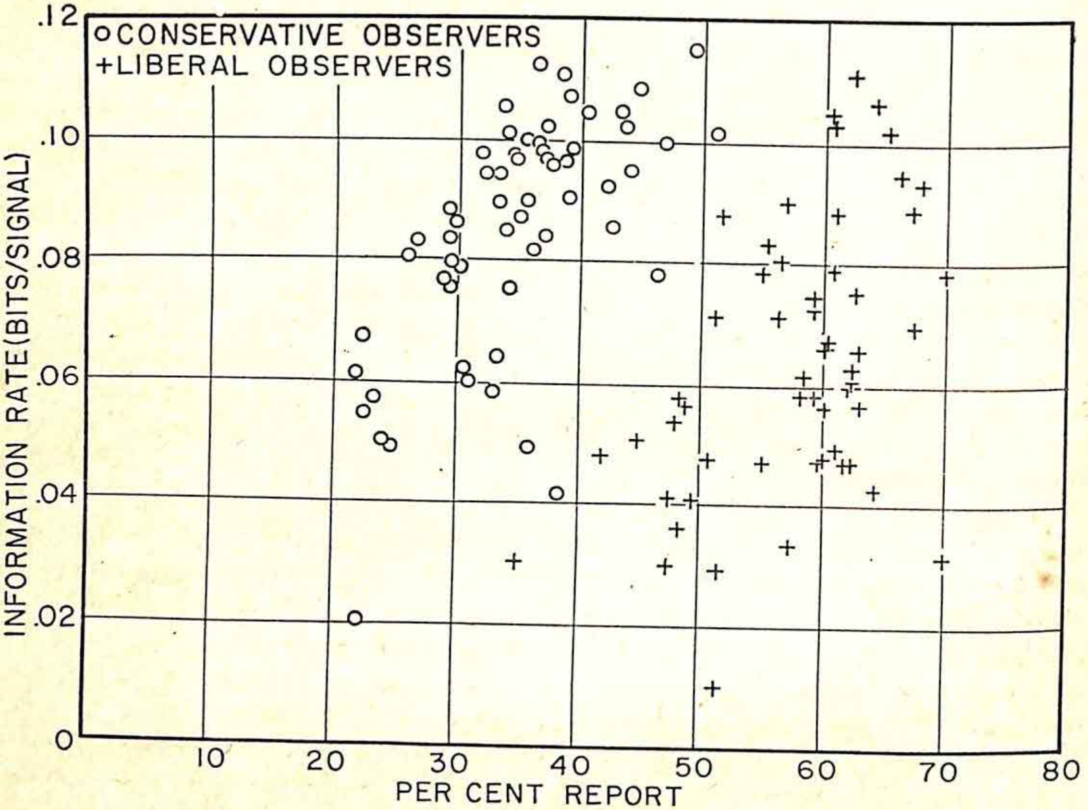


FIG. 7. Individual values of information transmitted per signal plotted against percentage report (signals and blanks).

question arises as to whether or not the increased report rate increases the amount of information being transmitted by way of the observer.

Shannon's (8) formula for rate of transmission in the presence of noise was applied to these data. [Rate $= H(x) + H(y) - H(x, y)$.] It should be noted that the absolute values of the calculated rates have very little meaning. These values depend on arbitrary factors, such as the percentage of signals (as opposed to blanks) used in the experiment, and the strength of signals used. The maximum rate of transmission under the conditions used (83.3 per cent signals and 16.7 per cent blanks) was .65 bits of information per symbol. The actual rates of transmission by the observers were far below this, since many of the signals were quite weak. However, the statistical characteristics and the strengths of the signals were the same under all experimental conditions, so that a comparison of conservative and liberal subjects seems justified.

For these calculations signals of all strengths were lumped together. A fourfold table was constructed for each observer, showing the percentage of times he reported a signal when there was and when there was not a signal actually given, and the corresponding percentages for his failures to report. The marginal totals on one side gave then the percentage of signals and blanks (the same for all observers), and on the other side the observer's percentage of reports. The formula given above was applied to these tables.

Figure 7 summarizes these calculations for the data from the main experiment. It shows rate of information in bits per signal plotted against the percentage of times the observer reported (to a signal or a blank). Each dot represents an observer working under the conservative attitude, and each cross a liberal observer.

Inspection of the plot shows that the liberal observers transmitted somewhat less information than did the conservative observers. The mean rates were .064 and .085 respectively.

On the other hand, it may be seen that within each group the information rate is positively correlated with the percentage of responses. Particularly in the case of the conservative observers, those with high report rates transmit more information. However, this correlation is due in part, if not in full, to differences in sensitivity of the observers. The men with high report rates are those who most effectively discriminated signals from noise. If it were possible to compare only observers equal in sensitivity, much of this correlation would probably disappear.

There is one fairly strong indication that this is the case. Although over-all report rate and false report rate are positively (but not linearly, see Fig. 8) correlated, there is no correlation between information and false report rate. Figure 9 shows this plot for liberal and conservative observers, and indicates that, within each group, the correlation is zero or slightly negative. This suggests that those observers in the conservative group who adopted a more liberal attitude (and hence increased both report rate and false report rate) did not tend to transmit much more information.

It may be that the observer who is too cautious in reporting signals trans-

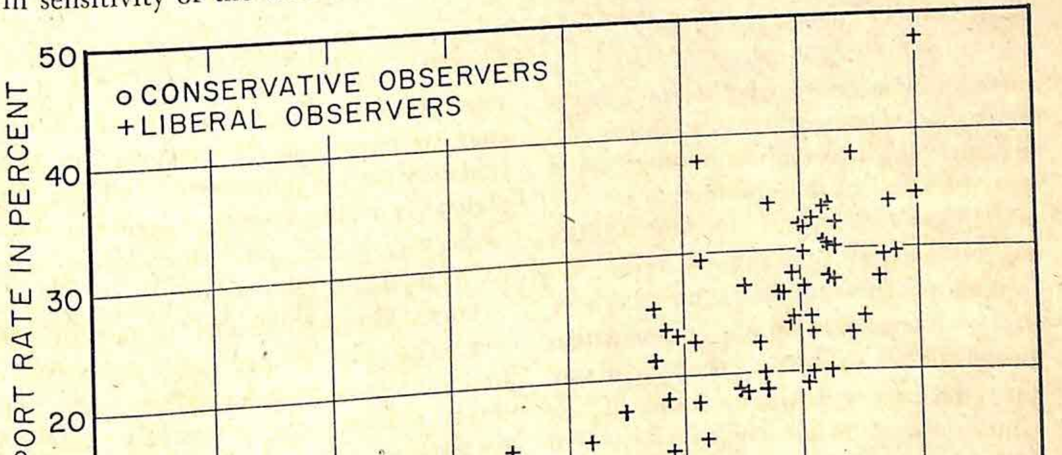


FIG. 8. Individual values of false report rate (percentage report of blanks) plotted against percentage report (signals and blanks).

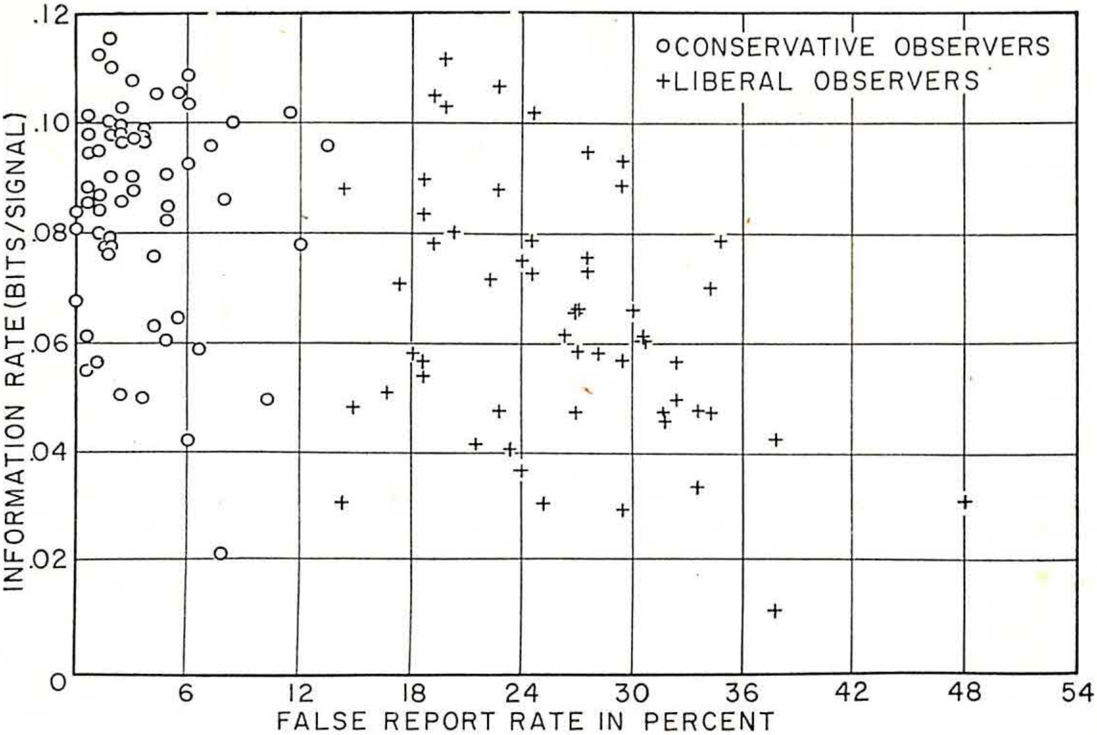


FIG. 9. Individual values of information transmitted per signal plotted against false report rate.

mits less than his maximum amount of information. These data give no direct indication of it. They do indicate, on the other hand, that when the attitude toward reporting becomes liberal, the rate of transmission of information drops.

E. SPECIAL GROUPS—FOUR CATEGORIES OF RESPONSE

Two special groups were tested to provide more information on the interaction of cumulative response curves and attitude toward reporting. With the exception of minor differences in the initial training procedures, these two groups of five men each were given the same experimental treatment. Both groups were tested over a five-day period, rather than the usual one-day period, so that they were relatively well practiced. The members of both groups showed every evidence of interest in the experiment, so that their results are judged to be not only more stable than those of the other groups, but also more nearly representative of what a cooperative and competent group of observers can do. These men were unselected with respect to auditory acuity, and seemed to be representative of the larger sample tested.

Instead of the usual yes-no categories of response, these observers were asked to make a four-category judgment in response to each signal. They were asked to indicate whether they were sure they heard the signal, they thought they heard the signal, they guessed that a signal had been presented although they couldn't hear it, or they guessed that no signal had been presented. The first category of response was intended to be equivalent to the responses of the conservative groups. A response falling in either the first or the second category was equivalent to a response by one of the liberal groups. The sum of responses in cate-

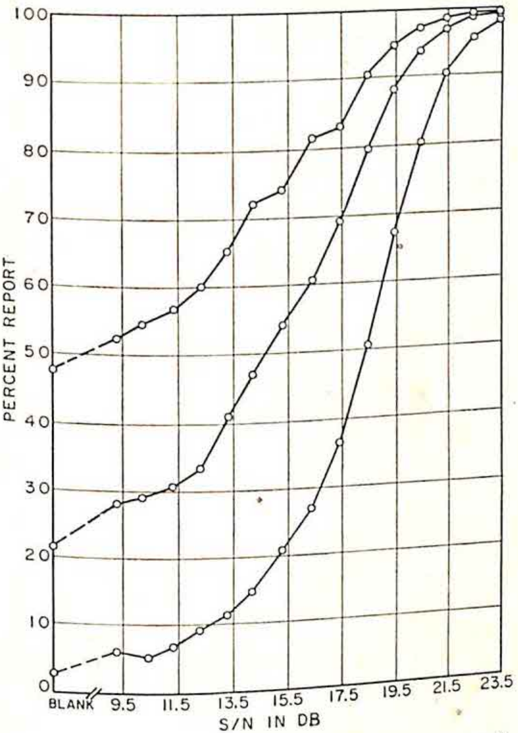


FIG. 10. Percentage report of observers responding in four categories. Bottom line is percentage "certain" responses (conservative attitude). Middle line is "certain" responses plus "hunches" (liberal attitude). Top line is first two plus "guesses" (radical attitude).

gories one, two, and three provided us with a new attitude toward reporting, which we have called the "radical" attitude.

Figure 10 shows a plot of the cumulative percentage of reports (average for ten observers) under each of these three attitudes. Although these curves are taken from responses to the same signals, the lower curves (for conservative and liberal attitudes) are quite like those for groups working under only one attitude. A comparison of Fig. 10 and 3 shows that the four-category groups were slightly superior to the average of all observers working under a single attitude. Their false report rates, under both liberal and conservative attitudes, were a little lower, but for most signal strengths, the percentage of reports was a little higher. This difference is probably attributable to the fact that these two groups were particularly highly motivated.

The table presenting individual data for each of the ten observers may be obtained from the American Documentation Institute.[2] Each man made from 130 to 160 judgments at each signal strength (three times that many on blanks) so the individual curves are fairly stable. The table shows individual differences. Observers 4 and 9 were consistently low in their report rates. Observer 7 was much too free in reporting under the conservative attitude (8.1 per cent false reports), although his liberal and radical curves are comparable to the rest of the group. The remaining seven observers are fairly homogeneous, and separate averages are given for this subgroup. The average of all 10 observers, corrected for guessing, is given. A comparison of the corrected averages for the liberal and radical attitudes shows that this additional category of judgment also adds something more than sheer guessing, although the addition in this case was not as large as in the shift from conservative to liberal attitudes.

TABLE 1

RATE OF INFORMATION TRANSMISSION IN BITS PER SIGNAL AND RESPONSE RATES FOR INDIVIDUAL OBSERVERS MAKING FOUR-CATEGORY JUDGMENTS

Observer	Rate of Information Transmission in Bits/Signal			Response Rate		
	Cons.	*Lib.*	*Rad.*	*Cons.*	*Lib.*	*Rad.*
1	.103	.098	.069	.369	.646	.788
2	.081	.072	.038	.320	.588	.840
3	.095	.072	.023	.370	.603	.806
4	.065	.074	.030	.254	.440	.654
5	.104	.077	.041	.362	.617	.812
6	.110	.070	.079	.388	.543	.684
7	.088	.081	.062	.441	.557	.676
8	.097	.078	.047	.374	.582	.717
9	.061	.043	.024	.281	.497	.665
10	.091	.092	.074	.331	.562	.712
Avg.	.090	.076	.049	.349	.564	.735

---

[2] For table of individual data for observers making four-category judgments, order Document 3932 from ADI Auxiliary Publications Project, Photoduplication Service, Library of Congress, Washington 25, D.C., remitting $1.25 for microfilm (images 1 inch high on standard 35 mm. motion picture film) or $1.25 for photoprint readable without optical aid.

Since the recording of four response categories took two of the five rows on the teletype tape, only two observers could be run simultaneously under this condition. Consequently, correlations between observers in these groups are based on too few observations to be at all reliable. They were computed, however, and were comparable to those reported in the section on intercorrelations of observers.

Information rates were also calculated for each of the ten observers under each of the three attitudes. The results are shown in Table 1. Again it may be seen that the average rate of information decreases as observers adopt a more liberal attitude toward reporting. Only two of the ten observers increased their rate of transmission in going from the conservative to the liberal attitude. One observer increased his rate in going from the liberal to the radical attitude.

IV. THE PROPOSED MODEL AND ITS APPLICATION

A. The Model

The statistical character of a sensory threshold has long been recognized, but only in terms of a distribution of sensitivity. That is, it is usually assumed that an observer either does or does not hear a signal, and that the variability in report comes from a variation in momentary sensitivity. The usual psychophysical experiment does not even test the observer to see if he will report when there is no objective signal. If the test for false reports is used, the resulting reports are treated simply as errors, to be kept to a minimum.

We have shown that the false report should be treated as something more than an error in response, since observers allowed to make false reports show a real gain in sensitivity to signals. Fig. 3 showed the average cumulative response curve, under both liberal and conservative instructions toward reporting, corrected for the guessing implied by false report rate. If nothing but guessing was involved in the false reports, these two curves should have come together. The correction assumes that the observer either does or does not hear the signal. He reports positively when he does, and also reports positively, with a certain probability, when he doesn't. This probability can be taken as equal to the false report rate, since the signal should never be heard under these conditions. The failure of this correction to bring the curves together indicates the inadequacy of these assumptions.

As an alternative to the usual two-dimensional treatment of threshold, we wish to propose a three-dimensional model. The dimensions are signal-to-noise ratio (or signal strength), subjective intensity of signal, and probability density of that subjective intensity. Figure 11 gives an oversimplified diagram of the model. It shows, for each signal-to-noise ratio, a probability distribution of subjective intensity. It is assumed in the model that such a distribution exists for each signal-to-noise ratio, including blanks, that each distribution is normal, and that all have the same variance. The means of the distributions are here assumed to decrease linearly with S/N in db, but to decrease nonlinearly below

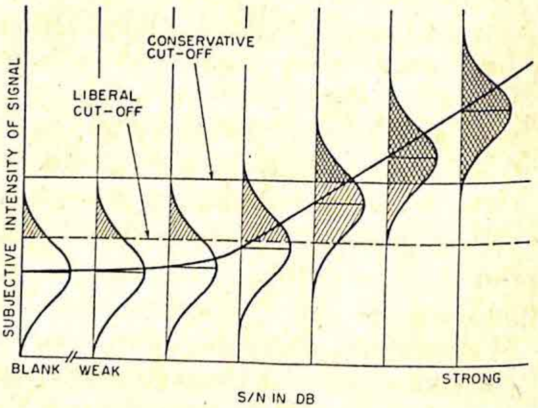


Fig. 11. Simplified model of the threshold.

some point. When the signals are very weak with respect to the noise, it matters little whether there is actually a signal or not. The blank (no signal) is assumed to be just the limiting case of weak signals.

The effect of attitude toward reporting is then to move a cut-off line along the dimension of subjective intensity. Any signal falling above this line will be reported, so that the cross-hatched areas represent the percentage of reports under conservative instructions, and the ruled areas are the corresponding percentages for liberal instructions.

Although these assumptions are not quite adequate, it is instructive to see how they compare with the usual assumptions. If we assume an all-or-none threshold, normally distributed with respect to signal-to-noise ratio in decibels, this model would be modified to the extent of straightening out the line relating the means of the normal distributions to S/N. Since the case of no signal represents a negatively infinite signal-to-noise ratio, this means that there should be no false reports. We found it impossible to prevent false reports completely although we could, by instruction, manipulate the rate at which they occurred. Moreover, all our plots of cumulative percentage of signals reported (even those "corrected for guessing") showed a deviation from normality in the lower end of the scale. It could be argued that this result could be more simply justified by changing the assumption about the physical scale against which we should expect normality of threshold distribution. If we used some transformation of S/N in db, we could normalize the lower branch of the distribution. However, no monotonic transformation will bring the blanks into line.

The main objection to this proposed model stems from its extreme flexibility. By changing the assumptions about the function relating the means of the distribution to S/N, or the assumptions about the relative variances of the distributions, or the assumptions about normality of distribution, it is possible to fit the model to almost anything. However, the two-dimensional model does not account adequately for our results, and if we can find a consistent set of assumptions for our model, it should prove useful.

1. Assumptions on Variability

Our first concern is with the variances of the distributions at different signal strengths. In the model as it is shown in Fig. 11, they are assumed to be constant. This assumption turns out to be inadequate, and needs to be investigated. It is convenient to divide the variance into two components, the variability common to observers and the individual variability.

a. Variance Common to All Observers

It was shown in the results section that observers listening to the same signal and the same noise tend to report at the same time and to fail to report at the same time. This correlation may be due in part to factors such as shifts in group morale, fatigue, and sequence of signals, but the principal source is the noise. Since all five observers are listening simultaneously, any random fluctuation in the noise will affect all of them to the same extent. The relative magnitude of this common variance is indicated by the size of the correlation between observers.

If we assume that subjective intensity is a random variable composed of the sum of noise intensity and individual variation, and also that individual variation is the same for two observers, then the correlation coefficient, r, will be the percentage of the total variance that is attributable to noise. The noise is always the same, so this is one feature of the situation that is independent of signal strength or of attitude. In the case of observers working under the conservative attitude, it is assumed that the value of r remains constant at .55 regardless of signal strength. It follows, then, that the total variance of the distribution of subjective intensity is independent of signal strength in this case.

When observers are working under the liberal attitude, however, the correlation between observers is a function of signal strength, being lower for weak signals than for strong signals. This means that the noise variance, which is independent of the effects of attitude, is now a smaller percentage of the total variance. This, in turn, shows that the other factor contributing to the total variance, the individual variance, is greater for weak signals than for strong.

b. Individual Variability

The sources of individual variability are not quite so easy to identify. There should be at least two, although finding objective evidence for either is quite difficult. There is variation in sensitivity and variation in cut-off level.

There is no direct evidence for variation in sensitivity, but it seems to be a theoretical possibility. Actually, it is probably not sensitivity to an 800-cycle tone that is varying, since such a variation should be accompanied by variation in sensitivity to the masking noise in the critical band. If this were the case, signal-to-noise level should remain unchanged, and the probability of a report should not change. However, it is reasonable to assume that the receptor mechanism itself contributes varying amounts of internal noise, which could change the signal-to-noise level.

In terms of our model, such variation in internal noise could contribute to the 45 per cent of the total variance that is associated with the individual working under conservative instructions. This part of the individual variance could not change with instructions, since the observer can make judgments with respect to both attitudes simultaneously.

The second source of individual variability, variation in cut-off level, is the one that must vary with instructions. The intercorrelations of observers show that individual variance on weak signals is greater under liberal than under conservative instructions. The external noise is independent of attitude, and so is internal noise, if it exists. This process of elimination leaves cut-off variability as the only source of a difference.

Variation in cut-off level has an effect that is easy to handle. The result of variation in cut-off level is mathematically the same as that of a stable cut-off level on a distribution whose variance is the sum of the variance of the original distribution and the variance of the cut-off distribution.

Given two normal distributions of x and y (representing here the distribution of subjective intensity and the distribution of cut-off level, respectively), with variances σ_x^2 and σ_y^2, we want to find the probability of x exceeding y. This probability will vary, of course, with the value of c. (c is the mean value of the cut-off distribution, expressed in units equal to th σ of the main distribution.)

We want $P(x > y) = P[(x - y) > 0]$. The distribution of $(x - y)$ will be normal (since x and y are both normal); with the mean of $M_x - M_y = -C$, and variance of $\sigma_x^2 + \sigma_y^2$ (x and y assumed to be independent). Thus, if we let $u = (x - y)$,

$$P[(x-y)>0]=\int_0^\infty \frac{1}{\sqrt{2\pi}\,\sigma_u}\, e^{-\frac{1}{2}\left(\frac{u+c}{\sigma_u}\right)^2} du$$

Let

$$u'=\frac{u+c}{\sigma_u},$$

and

$$P[(x-y)>0]=\int_{c/\sigma_u}^\infty \frac{1}{\sqrt{2\pi}}\, e^{-\frac{u'^2}{2}} du'$$

$$=\int_{c/\sqrt{\sigma_x^2+\sigma_y^2}}^\infty \frac{1}{\sqrt{2\pi}}\, e^{-\frac{u'^2}{2}} du$$

Thus the variable cut-off has the same effect as a constant cut-off equal to the mean of the cut-off distribution operating on a distribution of increased variance.

It would be very helpful if we could develop a theory of the change in cut-off variance as a function of signal strength. In this case we can deduce from the intercorrelations that it must change, and can even estimate how much it changes. However, in the case of an absolute threshold there could be little common variance, and hence no way to estimate changes in variance with changes in attitude or signal strength.

Since the observer does not know the actual signal strength at the time he makes his judgment, it is a little unreasonable that his variability is related to signal strength. We might assume that a change in attitude brings an increase in cut-off variability that is independent of actual signal strength, but which is applied to only that proportion of the judgments falling below the conservative cut-off level. This assumption would mean that most of the judgments on weak signals would be subject to the liberal variance, and most of those on strong signals to the conservative variance, so the trend in liberal intercorrelations as a function of signal strength would be in the right direction. However, the correlations that are estimated from these assumptions do not correspond very well with the observed values of r. Moreover, the resulting distribution is not normal, but consists of elements of two different normal distributions.

Since we are unable to justify anything more elaborate, we have made our assumptions as simple as possible. In terms of the model, the subjective intensity distributions under conservative instructions are assumed to have the same variance. This variance is arbitrarily taken as unity, with .55 associated with noise and .45 with individuals. For the liberal distributions, the variance is assumed to vary with signal strength. The size of the variance can be estimated from knowledge of the intercorrelations of observers. Since r is the percentage of common variance, and since the noise variance is .55, regardless of attitude,

$$r_{(\text{liberal})}=\frac{.55}{\sigma^2_{(\text{liberal})}}\,;\quad \text{or,}\quad \sigma_{(\text{liberal})}=\sqrt{\frac{.55}{r_{(\text{liberal})}}}\,.$$

2. *Other Assumptions*

One other assumption calls for reconsideration. It deals with the normality of distribution of the subjective intensity distributions. We have no real justification for this choice, and can suggest no empirical check. However, it seems to provide a good approximation, and we need to make only a minor change in the assumption.

Observers working over long testing sessions find it extremely difficult to keep constantly alert. It was pointed out above that in a few instances in this experiment observers were known to have fallen asleep for short periods. There were undoubtedly other instances, unknown to the experimenter, in which the observer lapsed into a state that might well have been sleep, even if his eyes were open. Better testing conditions and more careful selection of observers might have improved this situation, but it is a factor in any threshold experiment.

The effect of this lapse is to put the observer into another state, as far as probability of report is concerned. He may occasionally awake and, realizing that he has missed a signal, guess that one occurred. This might happen occasionally under the radical attitude, but more rarely under the liberal, and not at all under the conservative attitude. We assume, therefore, that the subjective intensity distribution is actually bimodal, with the larger part of the area (98 per cent or more) lying under a normal shaped curve. The remaining 2 per cent or less of the area falls far below, out of reach of any cut-off line. When most of the area lies below the cut-off line, this assumption will have negligible effect. However, for strong signals, the effect is more important.

B. APPLICATION OF THE MODEL

1. *Application to Average Individual Curves*

We have spoken throughout about the distribution of subjective intensity. This is a convenient fiction, designed to facilitate discussion. Our measurements are of percentage of report, and we assume only that there is some normally distributed variable underlying the report, and that the integral of this distribution above the cut-off level gives the percentage of reports.

To fit the model to the data we have fixed the means of the distributions with respect to a constant conservative cut-off level. We assume that all the conservative distributions have the same variance, and since we have no independent scale on which to measure this variance, we arbitrarily take it as unity. This allows us to enter the table of the integral of the normal curve, find the value of the integral equal to our per cent of report for a given signal, and read off the corresponding x/σ score. The mean of that distribution must be this number of units above (or below) the conservative cut-off line. The resulting plot of means has the same form as a plot of the percentages on normal probability paper.

Once the distribution means have been fixed, the liberal cut-off line can be plotted. Because of the assumed variability of the liberal cut-off line, the σ of the liberal distributions is no longer unity, but varies with signal strength. The value of the σ is obtained by computing $\sigma_L = \sqrt{.55/r}$, where r is the correlation between observers for a given signal strength. Rather than use the observed values of r, we drew a freehand smoothed curve through these values, and used the values from the smoothed curve. The observed points were quite variable, and the smoothed curve cannot be regarded with a great deal of confidence. It followed roughly the form of an ogive.

After the σ of the liberal distribution has been obtained, the distance from mean to liberal cut-off line can be gotten by looking up the x/σ value for the observed percentage of liberal reports, and multiplying by σ. If the model fits, all of these cut-off points should fall on a horizontal line.

One further correction is needed. Because of the occasional lapses of ob-

TABLE 2

COMPUTATIONS FOR FITTING THE MODEL TO DATA FROM THE MAIN EXPERIMENT

S/N	Conservative			Liberal			
	%Rep.	*Corr.*	*Z*	*% Rep.*	*Corr.*	*r*	*Z*
Blank	.035	.036	+1.80	.259	.264	.25	+ .93
9.5	.059	.060	+1.56	.306	.312	.34	+ .62
10.5	.055	.056	+1.59	.336	.342	.35	+ .51
11.5	.072	.073	+1.45	.361	.368	.36	+ .42
12.5	.088	.090	+1.34	.370	.377	.37	+ .38
13.5	.111	.113	+1.21	.438	.446	.38	+ .17
14.5	.145	.148	+1.05	.481	.490	.39	+ .04
15.5	.199	.203	+ .83	.538	.548	.41	− .14
16.5	.276	.281	+ .58	.609	.620	.42	− .35
17.5	.368	.375	+ .32	.699	.712	.44	− .63
18.5	.496	.505	− .01	.779	.793	.46	− .89
19.5	.659	.671	− .44	.868	.884	.47	−1.30
20.5	.806	.821	−. 92	.928	.945	.48	−1.71
21.5	.897	.913	−1.36	.961	.978	.49	−2.13
22.5	.945	.962	−1.77	.972	.9898	.50	−2.44
23.5	.971	.989	−2.29	.981	.99895	.50	−3.22

Note: *r* taken as .55 for conservative observers throughout.
The correction is for an assumed 1.8% response failures.

servers, it is assumed that the distribution is actually bimodal, with the main portion of the curve being normal in form, and a small part of the area way down on the scale, where no cut-off line could reach it. We want, therefore, to deal only with the main, upper curve, and want to adjust its area to unity. If 1.8 per cent of the area is assumed to fall in the smaller part of the curve, for example, then all areas should be multiplied by the reciprocal of .982 to give a corrected report.

The computations for the model-fitting to the data from the main experiment are given in Table 2. The correction factor applied to the report rates was 1/.982 or 1.0183. This assumes that 1.8 per cent of the judgments fell outside the distribution, which is probably too high a figure. This choice was arbitrary and was made to make the report of the strongest signal under the liberal instructions nearly unity. The corrected plot is shown in Fig. 12 and it may be seen that the liberal cut-off points fall close to a straight line, one sigma below the conservative cut-off. There is a slight tendency for the cut-off points to rise for the strong signals, despite the correction. The uncorrected data provide a plot that is, on the scale of this graph, indistinguishable from the corrected one below 18.5 db. Above this point, the uncorrected graph shows the means of the distributions sagging down toward the conservative line. For example, the dis-

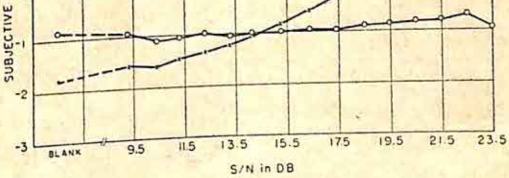


FIG. 12. Model applied to average data, all observers. For conservative observers, *r* taken as .55, independent of signal strength. For liberal observers *r* taken from smooth curve fitted to Fig. 6.

tribution mean for the strongest signal is is 2.29 corrected, and 2.90 uncorrected. The corrected liberal cut-off point is −3.22; uncorrected it is −2.17, or only .27 below the conservative cut-off, instead of the 1.00 that held for the weaker and middle-range signals.

2. *Application to Four-Category Judgments*

The same sorts of curves were plotted for the ten observers who used the four-category judgment. In this case, the area in the smaller, excluded distribution was taken as .006, instead of .018. This figure was chosen to make the percentage of report of the strongest signal under the radical attitude nearly one. The uncorrected percentage was .9935, and the corrected percentage .9995. Obviously, the percentages were not observed with this degree of accuracy, so that the correction is fairly arbitrary.

The correlations used were taken from the main experiment for the liberal curve. For the radical attitude, there were no correlational data. The correlations used were taken from the smoothed curve for liberal correlations, with the transformation $r_R = 2r_L - .55$. This transformation preserves the form of the curve relating liberal correlations to signal strength, and doubles its distance from the conservative correlation.

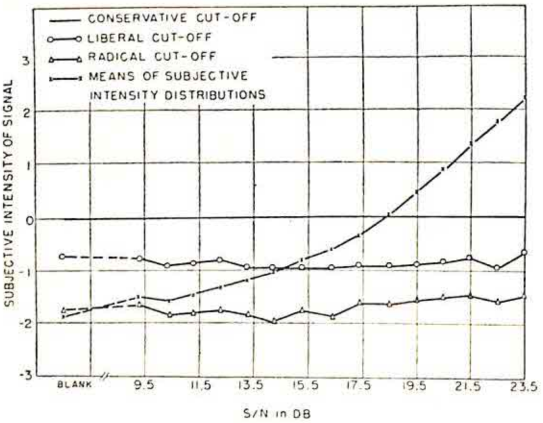


FIG. 13. Model applied to data from 10 observers making four-category judgments.

The resulting curves are plotted in Fig. 13. Again, the fit is fairly good for the liberal cut-off line, although there tends to be a rise in the line for the stronger signals. The radical cut-off line shows this effect to a much more marked extent. There are several possible reasons for this rather poor fit on the radical cut-off. Our estimates of the variances of the radical cut-offs were complete guesswork. We may have underestimated the variances of the distributions for stronger signals. It is also possible that we have underestimated these variances in the case of the liberal judgments as well. Those variances were estimated from the intercorrelations of observers, but as the signals got stronger, more of the correlations were indeterminant. This indicates the possibility of a systematic bias, since only the observers who failed to respond got into the correlations. This situation could be brought about by occasional signal failures, or by the fact that some observers were less sensitive to the signals and were, in effect, listening to signals a few db weaker than those presented to other observers.

Another possible reason for the failure of the fit of the radical cut-off line lies in our assumption of normality of the distributions of subjective intensity. This assumption fits fairly well in the middle range of the distribution, but it is quite possible that the assumption fails at the extremes. The actual form of the distribution need not deviate greatly from normality. It should be kept in mind that this method of presenting the data greatly magnifies deviations at the extreme of the distribution. In terms of deviation of observed percentage report from the expected value, the fit is not bad. For example, in Fig. 13 the average percentage report was .9935 instead of .9995 for the strongest signal under the radical attitude. If there had been only one of the 1550 signals missed, instead of 11, the radical cut-off would have fitted without correction. However, the systematic nature of the deviation as plotted would seem to indicate some faulty assumption.

3. Application to the Concept of Information Transmission

It was shown in the section on results that the conservative observers transmitted more information. There was some indication, however, that this statement is not universally true. The extreme of conservatism would be the reporting of no signals or blanks, whereas the extreme of liberalism would be a report at every opportunity—all signals and all blanks. Neither of these extreme modes of response would transmit any information, so that, logically, we must have at least one intermediate mode of response that transmits a maximum amount of information.

Application of the model proposed here makes it possible to plot rate of information transmission as a function of false report rate. It is assumed that as the cut-off lines are moved down, the variances of the individual distributions of subjective intensity increase in a systematic fashion. The values of these variances were estimated from interpolation or extrapolation of the observed values. For example, for any false report rate less than 3.5 per cent the intercorrelations between observers were assumed to be .6 for all signal strengths. For false report rates greater than 3.5 per cent the correlations were assumed to decrease toward the values obtained for the liberal observers. The expected report percentage of all signals was then computed for each false report rate and the information transmission rate was computed.

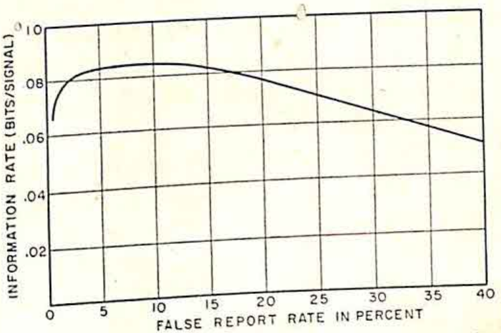


FIG. 14. Calculated (theoretical) values of information transmission plotted against false report rate.

The result is shown in Fig. 14. The graph shows that a single maximum exists, somewhere near 5 per cent false reports. Any false report rate less than 2 per cent leads to some loss of information, and suppression of false reports much below 1 per cent leads to serious loss.

4. Application to Multiple Observers

If each of the five individuals has a probability p of reporting a signal of a given strength, then the probability that at least one will report it is given by one minus the probability that all will fail to report. The joint probability that all will fail to report when the cut-off line is at a given value a, is the integral of the normal correlational surface up to the value a on each of the five variables. The probability that at least two (two or more) will report is this first probability minus the probability that four will be below the cut-off and one above. This latter probability is given by another segment of volume under the normal correlational surface. Similarly, the rest of the n-or-more curves may be obtained.

The method of integration of the normal correlational surface discussed in the introduction allows us to get all these values. The general procedure was that of summing, for all values of t in steps of .1, the product of the normal ordinate and the area to the fifth power from $t - a$ to infinity, for a normal distribution with $M = 0$, $\sigma = 1$. Schematically,

$$(\text{Vol. up to } a) = \frac{1}{\sigma_6} \sum_{t=-\infty}^{\infty} \left(\text{ord. at } \frac{t}{\sigma_6}\right) (\text{Area from } t-a \text{ to } \infty)^5.$$

By taking this sum for lower powers of the area, all the marginals of the five-variable surface can be computed. From the marginals it is possible to build up the volumes in the five dimensional surface for all possible splits. By changing the value of σ_6 ($\sigma_6 = \sqrt{r/\ (1\text{-}r)}$) it is possible to get the integral for any value of r. These sums must be computed for enough values of a to plot the cumulative distribution of the n-or-more criteria. For the case of the conservative observers this is relatively simple, since the correlation is the same for all values of a (r is independent of signal strength). For the liberal observers it was necessary to use different values of r for each value of a.

In the model we have proposed, the distributions of subjective intensity refer to individual performance. To get the percentage of times n-or-more individuals respond to a given signal we replace the individual curves with the proper group-distribution. For one-or-more observers out of five, for example, the new distributions will have means that are shifted upward, and will be skewed, with the tail running downward. The curve for all five will be symmetrical to the one-or-more, with mean shifted down, and tail running upward. The two-or-more and four-or-more will also be symmetrical to each other, but shifted less. The three-or-more distribution will have the same mean as the individual distributions, but will have a smaller variance. Thus, the criterion of three-or-more reporting should yield 50 per cent reports for the same signal strength that gives 50 per cent for the individual. However, the group should report fewer of the weak signals, and more of the stronger ones. Examination of the data shows this to be the case. The 50 per cent points for three-or-more and for individuals correspond as closely as the graphs can be read for both the conservative and the liberal observers.

Before discussing the curves so constructed, it is necessary to consider the effect of the correction for inattention we made in applying the model to individual performances. This correction assumed that the observers were effectively out of the system for a small percentage of the judgments (1.8 per cent in the main experiment). This left only .982 as the area under the main distribution, and the observed percentages were divided by .982 to make the area unity. This correction could also be considered as an adjustment for a slight non-normality of the distribution.

The effect of this assumption on the n-or-more curve is similar to its effect on individual observers. In the case of two observers, Fig. 15 illustrates the result. If the inattentiveness of the two observers is assumed to be independent, then the joint distribution will be as pictured, and the main surface will have a volume of $1 - 2 \times .018 = .964$ under it. The volume under this surface from a to ∞ on both axes is the computed value for the normal surface multiplied

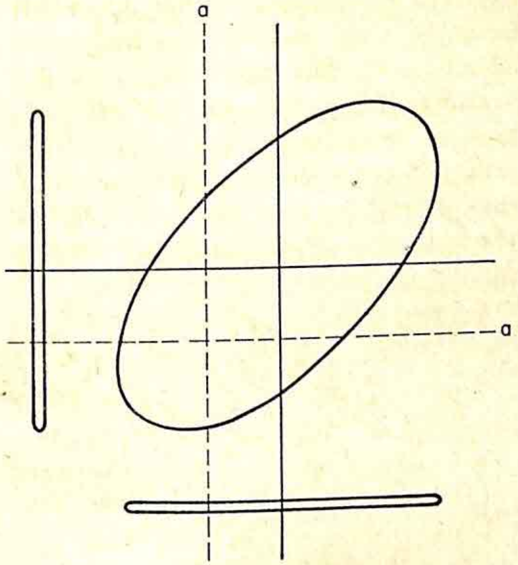


FIG. 15. Assumed effect of inattentiveness on joint distribution of two observers.

by .964. Similarly, for three variables, the correction consists of multiplying by $1 - 3 \times .018 = .946$ and so on. In other words, if each observer is in a state of nonresponsiveness part of the time, the number of times we can expect all observers to respond to the same signal will be decreased. If the individual never responds more than 98.2 per cent, then the group of five will never all respond to the signal more than 91.0 per cent. The corrected marginals allow computation of the cumulative distributions for all possible combinations of response.

This correction has very little effect except in the case of the curve for all-five-responding. This one is uniformly 91 per cent of its uncorrected height. However, this does not quite complete the account of the correction. These cumulative curves, corrected and uncorrected, are laid out on a scale with zero at the mean of the individual distribution, and units equal to the standard deviation of the individual. To get the expected uncorrected percentage report, we used the individual percentage report, converted this to an x/σ score, and read off the corresponding values from the uncorrected cumulative n-or-more curves. For the corrected percentage report for n-or-more observers, we used the corrected individual percentages (from Table 2). Thus the correction works in two ways, and the net effect is to increase very slightly the expected percentage report for all the curves except that for all-five-reporting. This one is lowered a little (3 per cent at the strongest signal) by the correction.

Figures 16 and 17 show, for the conservative and liberal observers, respectively, the corrected theoretical curves, with the observed values plotted on them. Because of the variation of r with signal strength in the case of the liberal observers, computation of the theoretical values was quite laborious and, in general, only

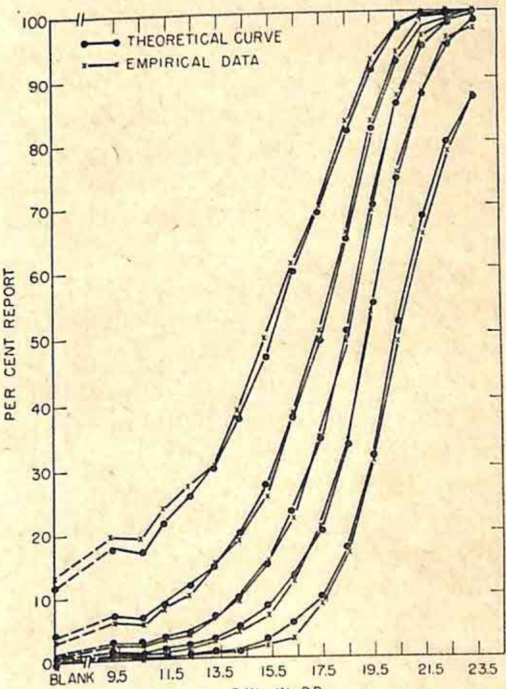


FIG. 16. Percentage report by n-or-more observers, conservative attitude. Solid points are average data from main experiment. Predicted values discussed in text.

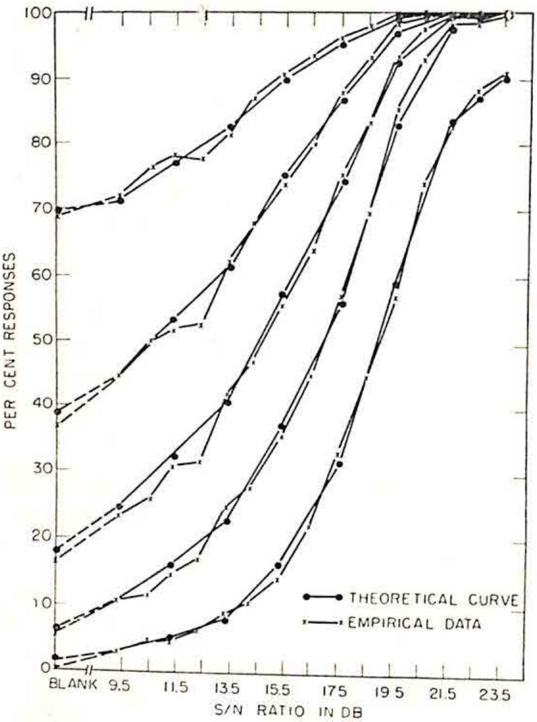


FIG. 17. Percentage report by n-or-more observers, liberal attitude. Data and predictions as in Fig. 16.

every other point is plotted.

There is a very high degree of agreement between expected and observed values. Some degree of agreement is to be expected, since the theoretical curves are based on the performance of the same individuals that made up the groups. However, we feel that the agreement helps to verify the general approach.

No new light is thrown on the validity of our model. Since the individual percentage report was used to predict the behavior of the group, we have used the cut-off lines shown in Fig. 12. However, had we straightened out the liberal cut-off line and used it, the difference in expected group performance would be changed so little as to be indistinguishable on the scale of our graphs.

5. *Extensions*

In terms of practical usefulness, the ideas explored here have been disappointing. Our results indicate that the liberal and conservative attitudes are very nearly equal in effectiveness when group criteria are selected to equate false report rates. It should be possible to deduce from the theory whether this is generally true, or only happens to be true for the special conditions of this experiment. Unfortunately, the extreme labor of computing the theoretical n-or-more curves makes such a deduction almost impossible. We can speculate, on the basis of the calculations of rate of information as a function of false report rate, that an extremely conservative attitude is inefficient, but over a fairly broad range of false report rates, attitude makes little difference.

The gain of the group over the individual is also disappointingly small. It is true that the conditions of this experiment were such as to minimize the gain of the group over the individual, but it is doubtful that conditions which maximize the gain will greatly increase it. The conditions for increasing the gain are those which increase individual variance, or which reduce correlation between individuals. That these conditions do not always produce a gain was shown in this experiment. Shifting from the conservative to the liberal attitude met both conditions, but did not increase the superiority of the group over the individual.

For five conservative observers, the criterion of two-or-more reporting will always give the best approximation to the individual false report rate, hence we can confine our attention to it. If the correlation between observers could go to zero while everything else remained unchanged, there would be a gain of roughly 1.5 times what it is now for the stronger signals. Referring to Fig. 5 we can see that this means about 1.5 db at the 50 per cent level, or 3 db at the 95 per cent level. Below the 25 per cent level there is actually a slight loss.

The factors which would tend to reduce correlation (slower rate of presentation of signals, variation in signal characteristics, etc.) should also tend to increase individual variability, and it would seem that the gain should be even greater. However, the interaction of the two factors is such that the gain does not increase.

Suppose that signals were presented less often. We would expect to find variability increased, presumably because of an increase in individual variability (variation in cut-off level). On the scale of our model (see Fig. 12) the total variance for conservative observers is one, with σ_c^2 (the common variance) = .55, and σ_I^2 (individual variance) = .45 for conservative observers. Suppose that the individual variances were quadrupled. It is difficult to make a realistic set of assumptions about this change in the cut-off variability that will leave it normally distributed, but we will assume that all cut-off scores are subject to the transformation $I' = 2I + 1.33$. This will increase the variance by 4 and increase the mean cut-off by 1.33. The mean increase is arbitrary, chosen to give 2 per cent false reports, since anything that increases the variability this much might be expected to reduce the over-all report rate. Our model would now have the same plot of mean subjective intensity, but the variance of each distribution would be $\sigma_c^2 + \sigma_I^2 = .55 + 1.80 = 2.35$ and $\sigma = 1.53$. The new cut-off line would be 3.14 units above the mean for the blanks. This yields an x/σ score of 2.05, and 2 per cent false reports. The expected percentage reports can then be obtained by reading from the graph (Fig. 12) the distance from the mean to the new cut-off line, and dividing each reading by 1.53 to get the new x/σ score. Entering the table of area under the normal distribution, we can get the expected percentage report. These figures, plotted against signal strength, will give a curve that rises less steeply than the present one, reaching only 72 per cent report for the strongest signal.

The correlation between observers under the assumed conditions would be $r = \sigma_c^2/(\sigma_c^2 + \sigma_I^2) = .55/2.35 = .23$. We can then estimate the expected percentage report for two-or-more observers, out of a group of five. The resulting curve shows a gain of about 1.5 db at the 50 per cent level, and 3.5 db at the 95 per cent level. Again the gain of the group over the individual is about 1.5 times the gain shown in our experimental results.

Similarly, we can estimate what should happen if individual variability remains as it was in our experiment, but all sources of common variance drop out. This, again, represents an extreme, rather than a reasonable, assumption. For an assumed false report rate of 3.5 per cent, the plot of expected report against signal strength now rises more steeply than in our experiment, reaching 95 per cent at about 18 db on the scale of our graphs. Here the steep slope of the curve leaves little room for improvement, and the expected gain of the group over the individual is approximately the same as it was in our results.

These calculations are rough, and are based on several very dubious assumptions, but they indicate that no gain of great practical significance can come from the use of a team of observers.

V. SUMMARY

This paper aimed at the presentation of an adequate theory of the multiple observer; that is, of the effectiveness of a group of observers working as a team. Before this could be done, it was necessary to develop a model of the threshold that takes false reports (i.e., positive reports when no signal is presented) into consideration.

The model proposed is three-dimensional, with a probability distribution of subjective intensities for each signal strength, including zero strength. Variation of the observer's attitude toward reporting moves a cut-off line up or down the dimension of subjective intensity. Any signal (or blank) exceeding this cut-off is reported. The necessity for such a model is indicated by conclusive evidence that the "guesses" made under the more liberal attitudes cannot be considered to be sheer guesses. If they were, a correction for guessing based on the false report rate should superimpose cumulative report curves, regardless of attitude. The data consistently failed, by a wide margin, to meet this assumption.

Two sets of experimental data were fitted to the model. In the main experiment 23 groups of five observers each listened for an 800-cycle tone of .41-sec. duration, against a background of broad-band noise. Twelve of these groups worked under instructions to be very conservative in reporting, and 11 groups were given more liberal instructions. The conservative groups were encouraged to avoid all false reports, but made an average of 3.5 per cent. The liberal groups were encouraged to guess whether or not they heard a signal, and had a false report rate of 25.9 per cent.

In a second experiment, ten observers were each given a large number of trials under instructions to make a four-category judgment on each tone. These observers made one response if they were sure they heard a tone, another if they thought they heard it, and for the remaining signals they guessed tone or blank.

The threshold model was applied by deducing variances of subjective intensity distributions from the intercorrelations of observers in the main experiment. The conservative cut-off line was taken as a reference point, and the cut-off points for other attitudes were plotted with respect to it. In both experiments the data fit the model fairly well, although there was some evidence of a discrepancy for the stronger signals.

With this model as a basis, a theory of the multiple observer was worked out and applied to the data from the main experiment. Agreement of theory and observation was satisfactory. Prospects for a practical application of the multiple observation unit were less satisfactory. Under the conditions of this experiment, the gain of the group over the individual was slight when false report rates were equated. The gain cannot be expressed adequately in terms of a single figure, since it was greater for stronger signals. However, at the 95 per cent report level, the group was only two db better than the individual.

REFERENCES

1. Hastings, C., Mosteller, F., Tukey, J. W., & Winsor, C. P. Low moments for small samples: a comparative study of order statistics. *Ann. math. Stat.*, 1947, **18**, 413-426.
2. Mood, A. M. *Introduction to the theory of statistics.* New York: McGraw-Hill, 1950.
3. National Defense Research Council, Div. 17, Sec. 3. Response characteristics of interphone equipment. O.S.R.D. Report No. 3105, 1944.
4. Pearson, K. Mathematical contributions to the theory of evolution. VII. On the correlation of characters not quantitatively measurable. *Philos. Trans. roy. Soc.*, Ser. A, 195, 1901, 1-47.
5. Pearson, K. *Tables for statisticians and biometricians.* Part II. London: Cambridge Univer. Press, 1931.
6. Schafer, T. H. Detection of a signal by several observers. Report No. 101, U. S. Naval Electronics Laboratory, San Diego, Calif., 1949.
7. Schafer, T. H. Influence of the preceding item in measurements of the noise-masked threshold by a modified constant method. *J. exp. Psychol.*, 1950, **40**, 365-371.
8. Shannon, C., & Weaver, W. *The mathematical theory of communication.* Urbana: Univer. of Ill. Press, 1949.

(Accepted for publication February 16, 1953)

Vol. 67, No. 10 Whole No. 360, 1953

Psychological Monographs: General and Applied

Passivity in a Class of Peptic Ulcer Patients[1]

Alvin Scodel
The Ohio State University

I. THE PROBLEM

THIS STUDY is an attempt to determine whether ulcer patients in psychotherapy exhibit personality trends measurably different from those of neurotic patients in psychotherapy. The research was suggested by previous reports in which ulcer patients, in contrast to normals or other psychosomatic patients, were said to have a specific psychological conflict or to comprise a special personality type (1, 7). An even more crucial problem, however, is whether the personality trends noted in the psychological investigations of ulcer patients are peculiar to them or characterize neurotic patients in general. If the latter situation obtains, there is little justification for speaking of an ulcer personality type or even a particular conflict in ulcer patients. It would be sufficient to label them as "neurotic." For example, a conflict over dependency needs is often described as the distinguishing feature of ulcer patients. Yet, if the same conflict is present to the same degree in neurotic patients, it obviously has no uniqueness for the ulcer group.

The present study concerns itself with a comparison of ulcer and neurotic patients in psychotherapy at a Veterans Administration mental hygiene clinic. Although it was not entirely possible to equate the groups for severity of psychopathology, a general factor of neuroticism in both groups was obtained by selecting patients on the basis of voluntary admission into psychotherapy. The ulcer sample is hypothesized to be within that group designated as "attitudinally counteractive and overtly passive," and is described in the next section.

II. THE ULCER PATIENT SYNDROME[2]

The writer hypothesizes an active-passive dichotomy in male ulcer patients and assigns a different etiological basis to each group. The active group would be selected from an upper-middle or upper class background where emphasis is placed on that kind of behavior which insures success in our culture. This group may be tentatively described as "attitudinally counteractive and overtly active." These patients refuse to accept their dependency needs, and their actual behavior is in conformity with their perception of themselves as active, efficient, and masculine. However, ego aspirations are set so high that the individual can never achieve them. He is forever falling short and his futile attempts at mastery serve only to increase his unconscious dependency needs.

---

[1] This monograph is based on a thesis submitted to the University of California in partial fulfillment of the degree of Doctor of Philosophy. It has been reviewed by the responsible officers of the Veterans Administration and is published with their approval. The opinions and conclusions are the writer's and do not necessarily reflect the opinion or policy of the Veterans Administration.

The writer wishes to express his indebtedness to Drs. Robert E. Harris, Audrey Schumacher, and Robert C. Tryon for their invaluable criticisms and suggestions during the planning and execution of this study. Special thanks are due to Dr. Jack Block and Mr. Dean Allen for their cooperation and encouragement throughout the duration of the study.

The writer is also grateful to staff members of the San Francisco Mental Hygiene Clinic for so graciously referring their patients to serve as subjects and to the patients themselves, who, in many instances, gave up several hours from their employment to participate in the study.

[2] A review of the literature containing previous theoretical discussions of the ulcer patient syndrome and a detailed elaboration of the writer's views have been given elsewhere (8).

The passive group is derived from a lower or lower-middle class background which fosters and encourages satisfaction with nonstriving occupational pursuits. This group, assumed to include the ulcer sample in the present research, may be labelled "attitudinally counteractive and overtly passive." These ulcer patients, like those in the active group, also deny their dependence, but their overt behavior is inhibited, conforming, and unassertive in contrast to their perception of themselves as active and productive persons. Such a self-evaluation serves as a defense against the culturally unacceptable dependency needs, and as long as the individual is able to obtain need gratification, the pretext of independent masculinity is maintained without any appreciable threat to the self. The individual's culturally approved perception of himself as masculine can remain inviolate while repressed passive and feminine needs are being gratified by an external source. Once the source of dependency is removed, however, repression is no longer an adequate defense. The environmentally induced frustration brings the dependence-independence conflict into focus since the conscious perception of self is directly attacked by reality. The person is faced with the impossible task of performing as if he were independent. His conscious phantasy of being efficient and masculine loses validity in the face of a self-imposed inability to act independently, on the one hand, and an environment which frustrates dependency, on the other. This insoluble situation increases the unconscious longing for dependence which, being incompatible with conscious perception, sets in operation those physiological processes that produce the ulcer.

III. SUBJECTS

A total of 72 subjects, all male, was used in the present study. Subjects were selected from the patient population of the San Francisco Mental Hygiene Clinic. There were 38 patients in the control group and 34 patients in the ulcer group. However, not all patients completed the battery of measures used in the study. Completion of the entire battery would ordinarily entail more than one session, and many patients would drop out of treatment, permanently or temporarily, before the battery could be completed.

A. *Diagnosis*. All patients in the ulcer group had a medical diagnosis of an ulcer which had been confirmed by X-ray studies. However, for inclusion in the study it was not required that the ulcer be active at the time the patient was seen for the study. Patients were included if their Veterans Administration records indicated that they had had an active ulcer which had been substantiated by X-ray diagnosis at any time within the five previous years. This criterion was designed to exclude any diagnoses which had been made prior to their entry into the service. In such cases no records were available, and the validity of the diagnosis would have had to rest on the personal testimony of the patient. The records listed all the cases but two as duodenal ulcer. The remaining two were diagnosed as gastric ulcer.

The control group was spread over a wide range of nosological categories. The diagnoses included anxiety states, conversion hysteria, and obsessional neuroses. Other cases were given less definitive diagnoses such as "dependency reaction" or "somatization reaction." Patients were not included in the control group if they carried a psychosomatic diagnosis or had reported in the intake interview a history of a psychosomatic illness. "Nonpsychosomatic" was defined as the absence, both at the time of testing and in the patient's history, of the following diseases: peptic ulcer, asthma, ulcerative colitis, rheumatoid arthritis, essential hypertension, and neurodermatitis. In view of Wolf and Wolff's assertion that an ulcer and other stomach complaints differ only in degree (11), patients with stomach complaints were also excluded from the control group. Patients were accepted if they had illnesses other than those just listed or if they had physical complaints of functional origin other than stomach complaints.

B. *Severity of psychopathology*. No patients in either the ulcer or control group displayed psychotic behavior at the time they were seen. The exclusion of a psychotic diagnosis was based on the provisional diagnosis obtained from the intake psychiatric interview or, if the patient was already in therapy at the time the study was initiated, from the patient's therapist.

C. *Length of time in psychotherapy*. Nineteen members of the ulcer group, or 56 per cent, and 17 members of the control group, or 45 per cent, were seen within the first month after they had gone through intake. The remaining patients had already been in therapy for a longer period at the time they were seen.

D. *Ethnic origin*. There were two Negroes in each group. The ulcer group contained three foreign-born patients (two Europeans, one South American), and the control group had two patients of European birth. The remaining 63 patients were native-born whites.

E. *Intelligence*. The vocabulary subtest of the Wechsler-Bellevue Intelligence Scale was administered to all 72 patients. In view of the correlation of .85 of this subtest with the total scale (10, p. 101), it was selected as an expedient measure of the patients' intellectual level. The means of the weighted scores were 12.4 for the ulcer group and 12.8 for the control group. This difference was not significant.

F. *Age*. The age range in the ulcer group was 23 to 50 years with a mean age of 35.4. In the control group the ages ranged from 23 to 53 years with a mean age of 31.2. Although the difference in mean age was significant at the .01 level, correlations between age and performance on each of the experimental measures were not significant.

G. *Socioeconomic status*. Although the lack of detailed information on many patients precluded a comparison of the socioeconomic background of the two groups, it can safely be stated that the majority of patients were of either upper-lower or lower-middle class membership. A study of consecutive admissions of San Francisco Mental Hygiene Clinic patients, undertaken one year before the present study was initiated, revealed that only 7 per cent of the patients entering therapy could be classified as having professional, managerial, or entrepreneurial status. It is principally for this reason that ulcer patients in the present sample are assumed to be in the passive rather than the active group. There is the additional reason that voluntary acceptance of psychotherapy in the Veterans Administration presupposes some degree of overt passivity. Payment is not required as it is in ordinary psychiatric practice, where the willingness to pay a fee is presumptive evidence that the patient feels motivated to change his status. Moreover, many of these patients are referred from other departments in the Veterans Administration or from public agencies, so that entrance into therapy is often a passive compliance with the suggestions of persons in authority.

IV. PROCEDURES

Three experiments were utilized, each designed to elicit different manifestations of overt passivity.

A. *Level-of-aspiration experiment*. The pursuit rotor was selected as the apparatus to be used in studying the patients' level of aspiration because it appeared to satisfy those criteria which have been described by most investigators as necessary for a good level-of-aspiration experiment. The pursuit rotor consists of an electrically driven turntable which is attached to an electrical clock. On the turntable there is a small brass disc. A brass stylus is also attached to the apparatus. Whenever the brass stylus makes contact with the disc, the length of time in which the two are in contact is recorded on the clock. The turntable can be adjusted to rotate at varying speeds. In the present experiment it rotated at a speed of 65 revolutions per minute.

Each patient was given a half-minute's practice on the apparatus, and the timing device was also explained to him. The patient was instructed to keep the stylus in contact with the disc as much as he could during each 20-second trial. Before each trial he was asked to predict his performance. At the conclusion of a trial he was told what his performance was, and was then asked to predict his next performance. This procedure was continued for a total of 11 trials, the first of which was ignored in the analysis of the data.

B. *Experiment on completed versus interrupted tasks.* The experimental material consisted of 16 pencil-and-paper tasks. These were selected for their general interest value and could be solved accurately by all subjects of average or better than average intelligence without being so easy as to destroy motivation. The problems included such tasks as working out a simple code, writing a paragraph backwards, selecting appropriate conclusions for logical syllogisms, rearranging jumbled sentences, etc. Before the experiment began, the experimenter said, "I am going to give you a series of tasks. I want you to do each one as quickly and accurately as you can." No other preliminary instructions were given. The patient was allowed to complete half of the tasks and was interrupted on the others when he was about three-quarters finished. If he was spending a disproportionate amount of time on the task, he was interrupted sooner because it seemed desirable to equate the lengths of time for the various tasks as much as possible. In the event that a patient was experiencing difficulty with a task which was to be completed, the experimenter removed the paper from him, and it was counted as unfinished. He was then allowed to complete the next task which was to have been interrupted.

Although the order of completion and interruption was constant, the order of presentation of the tasks was randomized for each patient. Torrey (9) has found that tasks which form a good Gestalt in the sense that a definite pattern has to be followed are recalled more readily after interruption than those tasks which can be broken down into discrete units. Since the tasks used in the present experiment varied greatly with respect to whether they were or were not good Gestalten, any bias which operated in favor of the recall of certain tasks was removed by randomization.

Recall was instituted three minutes after completion of the last task. During this time the experimenter pretended to write notes. At the end of three minutes the experimenter looked up and asked, as casually as possible, "I wonder if you would tell me what these problems were that you just worked on." The experimenter then recorded all those tasks that the patient enumerated. This procedure continued until a one-minute interval was reached which produced no responses. At no time did the experimenter imply that recall of the tasks was of any importance.

C. *Experiment on perseveration of inadequate hypotheses.* The procedure in this experiment was substantially the same as that used by Humphreys (4) in a study of the effects of the distribution of reinforcement upon the acquisition and extinction of conditioned responses. The apparatus consisted of two light bulbs juxtaposed on a vertical board. Behind the board were switches for each of the light bulbs. The experimenter sat behind the board and manipulated the switches. His movements were not observed by the patient who sat facing the board.

The following instructions were given to the patient:

"On each trial the light on your left will be turned on. Following that light, I may or may not turn on the light on your right. After I turn on the light on your left, and you think I am going to turn on the light on your right, put down an X after the number of the trial. If you think I am not going to turn on the light on your right, put down an O. After I turn on the light on your left, you will be given only a few seconds to make your choice so don't hesitate in writing down the X or O. However, do your best to catch on to the plan which I am going to follow."

The patient was then given 24 trials. On 12 of these trials the right light followed the left light. The right light was turned on in a random fashion so that the problem was actually insoluble. Following the twenty-fourth trial and without any break in the continuity of the presentations, the second light was never turned on. This procedure was continued until the patient had marked down 10 consecutive O's. This was the criterion for solution of the problem.

Despite the insolubility of the problem, the manner of presentation demonstrated to each patient that sometimes the second light went on and sometimes it did not. Some patients may have been operating on an erroneous hypothesis regarding the sequence of the reinforcements, others may have been following no specific hypothesis or hypotheses, but all patients can be said to have known that the second light came on periodically, if not in any systematic fashion.

D. *The Minnesota Multiphasic Personality Inventory (MMPI).* The present group of ulcer patients in psychotherapy was said to be attitudinally counteractive and overtly passive. For the most part, these patients perceive themselves as active, masculine, and efficient, despite their overtly passive behavior. The study of Ruesch and his associates (7) elicited clinical evidence suggesting that most ulcer patients' attitudes are characterized by extreme social conformity, a reluctance to express aggression openly, and an exaggerated lower-middle class value system. In order to test these statements some measure of attitudes was necessary which would reveal the discrepancy between the ulcer patient's actual behavior and his perception of himself. For this purpose the MMPI seemed ideally suited. The test provides both clinical scales and a large number of diverse items that are related to these predictions. For scoring and interpretation of the MMPI scales the reader is referred to the manual (3).

V. HYPOTHESES AND RESULTS

A. *Level-of-aspiration experiment.* This experiment assumes that a passive person will reveal his lack of mastery over his environment by setting low goals for himself in a situation where he is allowed a freedom of choice. A person may profess to be entirely self-confident and unafraid of failure, but in a situation which allows a freedom of choice in selecting goals, he may evade the possibility of failure by restricting himself to minimal achievements.

The patient's atypical responses on the pursuit rotor were used as a measure of his level of aspiration. Atypical responses included the following four kinds of responses:

1. Raising one's aspiration level after failure.
2. Maintaining one's aspiration level after failure.
3. Lowering one's aspiration level after success.
4. Maintaining one's aspiration level after success.

The sum of the first and second types of response comprised the patient's "high" responses, and of the third and fourth his "low" responses. The patient's level-of-aspiration score was obtained by algebraically summing the number of his high and low responses. For example, if he had three high responses and one low response, he received a score of 2; if he had one high response and two low responses, his score would be −1.

The hypothesis in this experiment was that the ulcer group would have a lower level of aspiration than the control group when the level of aspiration was measured by the difference between the number of atypical high responses and the number of atypical low responses.

Statistical comparison of the ulcer group ($N = 33$, Mean $= .67$, $SD = 2.61$) and the control group ($N = 36$, Mean $= 1.81$, $SD = 3.04$) yielded a t of 1.66, which is significant at the .05 level of confidence.[3]

The mean number of atypical responses for the two groups was almost identical. For the ulcer group the mean number was 3.70 and for the control group it was 3.65.

B. *Experiment on memory for completed versus interrupted tasks.* The rationale for this experiment was based on the assumption that, in Rosenzweig's terms (6), the ulcer patient in psychotherapy would be ego defensive rather than need persistive. Ego defensiveness, as such, cannot be equated with passivity, but the ego-defensive person manifests a low tolerance for failure by ignoring or rejecting his failures. The more active or need-persistive person attempts to overcome failure. In this somewhat derivative sense the ulcer patient's behavior in this experiment is construed as another manifestation of his overt passivity.

The present experiment takes as its starting point the same kind of informal (nonstressful) situation used by Zeigarnik (12), but makes an opposite prediction on the basis that a low tolerance for failure in ulcer patients would produce a greater recall of completed tasks. Since passivity also characterizes patients in the control group, the normal Zeigarnik effect cannot be expected to hold for many patients in this group. The hypothesis in this experiment, therefore, is that the reversal of the Zeigarnik effect will be more true of the ulcer than of the control group. There is an explicit assumption that if normals were tested in the same situation, selective recall would operate in favor of the incompleted tasks. Rosenzweig's experiment, as well as the original Zeigarnik study, can be cited as supporting evidence for this assumption.

---

[3] In view of the specific predictions, a one-tailed test of significance was used for comparisons of the groups.

The procedure used to test the prediction that the recall of completed tasks would be greater for the ulcer group employed recall differences in which the number of interrupted tasks recalled by each patient was subtracted from the number of completed tasks he had recalled. Statistical comparison of the ulcer group ($N = 29$, Mean $= .97$, $SD = 1.77$) and the control group ($N = 30$, Mean $= -.20$, $SD = 1.64$) yielded a t value of 2.60, which is significant at the .01 level of confidence.

Alternative statistical treatments, one employing the ratio of completed to interrupted tasks recalled, and the other the numbers of patients who recalled more completed or interrupted tasks, led to similar results.

The mean number of total tasks recalled was 8.4 for the ulcer group and 8.8 for the control group. This difference is not significant.

The writer prefers the term "selective recall" to repression in evaluating the results. Repression in the Freudian sense is usually construed to mean the purposeful exclusion of instinctual demands from awareness (2, p.148). The present experiment was intended only to test the low tolerance for failure in ulcer patients.

C. *Experiment on perseveration of inadequate hypotheses.* This experiment was designed to measure manifestations of overt passivity in a problem-solving situation. The experiment rests on two assumptions. The first assumption is that one of the correlates of passivity should be a delay in adapting to a sudden change in the external situation. This assumption led to the following hypothesis for the Humphreys problem: Given a situation in which set is firmly established and given an objective change in the external situation which makes the initial set no longer applicable, the ulcer group will perseverate longer on the initial set than the control group. Stated another way, the ulcer group will perseverate longer on an inadequate hypothesis before restructuring the situation than will the control group.

The second assumption in this experiment is that the passive person will continue to give a response when that response is reinforced by being proved correct. The less passive person will cast about for alternative solutions even after he has given a correct response. The second hypothesis in this experiment, then, was as follows: When the initial hypothesis has been perceived as incorrect and a new kind of response has been initiated which is applicable to the new situation, the ulcer group will maintain the new response more persistently than the control group. It was predicted that the ulcer patients would restructure the situation later, but, after they had restructured the situation, would have fewer erroneous hypotheses than the control group. An erroneous hypothesis in this instance was meant to be an unnecessary alternate hypothesis, since a new response that was applicable to the new situation had already been established.

In reference to the first prediction, "restructuring" was defined as a discernible change in the pattern of X's and O's following the twenty-fourth trial. In order to arrive at a measure of the point of restructuring the problem, the protocols were rated independently by five advanced graduate students in clinical psychology. Each rater was instructed to compare the pattern of the first 24 trials with the pattern of succeeding trials and to designate the number of the trial at which a definite transition appeared. Since a restructuring of the problem would often cover several or many trials, the raters were instructed to mark the change at the beginning of the transition. In this manner all raters employed the same criterion for change. A typical protocol will serve to clarify the rating procedure.

Trial
Number: 1 2 3 4 5 6 7 8 9 10 11 12
Response: X O X X O O X X X O X O
13 14 15 16 17 18 19 20 21 22 23 24
X X O O O X X O X O X X
25 26 27 28 29 30 31 32 33 34 35 36
O X O O X X X X X X X O
37 38 39 40 41 42 43 44 45
O O O O O O O O O

All raters assigned this patient a score of 29. Some patients obtained scores of 25, 26, or 27 although 28 was the first trial at which a change could have logically occurred. In the first 24 trials the light was never turned on or kept off more than twice in succession. Therefore, it would take at least three trials after 24 to perceive that the external situation had been altered. Scores below 28 are artifacts produced by the rating method, but it was deemed desirable to insure uniformity in the method chosen to evaluate change.

The tetrachoric correlations among raters ranged from .80 to .95 with a median correlation of .89. The composite rating has an estimated reliability of .97, which is the expected correlation between the obtained ratings and the ratings of five comparable judges.

The median of the five raters' scores for each patient was taken to be the patient's actual score. The positive skewness of the distributions limited the use of the conventional *t* test. As an alternative procedure, the pooled distributions were dichotomized at their median, which was 31, and a subsequent test of significance by chi square was made. A fourfold table of the distributions is presented in Table 1.

TABLE 1

POINTS OF RESTRUCTURING FOR THE ULCER AND CONTROL GROUPS

Group	Below Median	At Median and Above Median	Total
Ulcer	11	22	33
Control	23	15	38
Total	34	37	71

The obtained chi square of 5.23 is significant at the .05 level of confidence.

The prediction that the ulcer group, after restructuring the problem, would have fewer alternative (erroneous) solutions before achieving the correct solution was not confirmed. The method chosen to evaluate the number of erroneous hypotheses is described in detail in the writer's dissertation (8).

It is interesting to note that the groups were very similar with respect to the number of trials required for solution of the problem. The means for the ulcer and control groups were 42.70 and 40.76, respectively.[4] If the number of trials necessary for solution is assumed to be the same for the two groups, it may not be readily apparent why there is no significant difference in the number of erroneous hypotheses if the ulcer group restructures the problem later. The reason is that the number of trials is not necessarily an accurate index of the number of hypotheses. Consider the following two patterns.

1. O O O O O O X O O O O O O O X
2. O O X O O O X O O O O O X

In the first instance there are two hypotheses whereas there are three hypotheses in the second case, despite a smaller number of trials.

D. *Minnesota Multiphasic Personality*

---

[4] The number of trials required for solution was the number of the X which preceded the 10 consecutive O's.

Inventory. The experiments just described were designed, in general, to elicit various manifestations of overt passivity. The MMPI was selected to test the hypothesis that ulcer patients in psychotherapy exhibit attitudinal counteraction. That is, statements of the ulcer patient's attitudes toward himself and the world can be utilized to demonstrate his denial of passivity as well as the discrepancy between his evaluation of himself and his actual behavior. Moreover, it was expected that the test could be used as a check on the validity of Ruesch's statements concerning the personality traits of ulcer patients in general.

Thirty MMPI's were obtained from each group. The profile and item analyses will be treated separately.

1. *Profile analysis*. One of the conclusions in the Ruesch study was the evidence of a capacity for repression in ulcer patients. The same study also noted that these patients cannot express aggression openly, and use social techniques for the express purpose of obtaining approval from others. All these characteristics typify the syndrome of hysteria.

The *first hypothesis* was that the ulcer group would be higher on the Hysteria (*Hy*) scale than the control group.

The Ruesch study stressed the emphasis on socially proper behavior in ulcer patients. Although the Psychopathic Deviate (*Pd*) scale on the MMPI is usually used to infer a certain disregard for social norms (by virtue of a high score on the scale), a low score on this scale can be interpreted as a high degree of social conformity.

The *second hypothesis* was that the ulcer group would be lower on the *Pd* scale than the control group.

Another conclusion of the Ruesch study was that ulcer patients adhere to an exaggerated lower-middle class value system. Such values would encompass approval of those interests and activities which are regarded in our culture as masculine and, conversely, would be nonsympathetic to those pursuits which have a feminine connotation (e.g., aesthetic interests).

The *third hypothesis* was that the ulcer group would be lower on the Masculinity-Femininity (*Mf*) scale than the control group.

The means and sigmas of the scores on all the test scales are presented in Table 2. All scores were *K* corrected. For the interpretation of *K*, the reader is referred to the article by Meehl and Hathaway (5).

The differences between the *Hy* and *Mf* means are both significant at the .01 level of confidence. The difference between the *Pd* means, although in the predicted direction, is not significant. None of the other scales, with the exception of Hypochondriasis (*Hs*), showed significant differences between the groups. The difference in the *Hs* means probably has no interpretive value for the present study. Patients with psychosomatic disorders were ex-

TABLE 2

MEANS AND SIGMAS OF SCORES ON THE MMPI SCALES FOR THE ULCER AND CONTROL GROUPS

MMPI Scale	Mean		Sigma		*t*
	Ulcer	Control	Ulcer	Control	
Hs	78.50	65.70	18.34	17.37	2.78
D	78.17	79.80	15.57	15.39	0.41
Hy	73.33	66.00	12.83	10.61	2.42
Pd	65.07	67.83	11.66	13.53	0.85
Mf	63.30	70.30	8.26	13.22	2.46
Pa	58.23	59.30	10.89	10.75	0.38
Pt	73.83	72.06	15.80	13.44	0.47
Sc	69.33	72.40	16.32	16.81	0.72
Ma	58.47	58.00	10.30	10.28	0.18

cluded from the control group and, for this reason alone, one would expect a smaller number of somatic complaints in the control group.

2 *Item analysis.* An item analysis was done on all of the 550 MMPI items. It was predicted that differentiating items would reveal the following cluster of attitudes to be typical of ulcer patients.

a. Psychological inadequacies are rigorously denied.

b. Femininity is denied and masculinity is exaggerated.

c. Masochistic, self-depreciating attitudes are denied.

d. Dependence is denied, but, at the same time, there is a disinclination toward self-assertion. One must be capable, but must not proclaim his individuality.

e. There is an emphasis on proper and conventional behavior in order to avoid censure from others.

f. An impunitive rather than intropunitive or extrapunitive attitude toward aggression exists. However, aggression is expressed when such aggression would be socially approved or would lead to a feeling of moral self-righteousness.

g. Moralistic and puritanical attitudes toward sex and religion prevail.

h. There is an extroverted, socially oriented "philosophy of life" in contrast to an introverted, introspective, ruminative attitude.

The first four attitudes refer to the ulcer patient's attitudinal counteraction. In general, these attitudes deal with the ulcer patient's perception of himself as active, efficient, and masculine. The denial of dependence had to be qualified because of those predictions that related to the ulcer patient's desire for social approval. If he manifested his denial of dependence by uninhibited self-assertion, he would be inviting social disapproval. The last four attitudes relate to the ulcer patient's exaggerated social orientation and his emphasis on cultural proprieties.

After the number of "true" and "false" responses to each item had been tabulated for the two groups, those items were selected which differentiated between the two groups at the .10 level of confidence. Of the 74 items which were selected in this manner, 44 were also significant at the .05 level.[5] The lenient .10 level of confidence was used in order to cull more items that would tap the characteristic attitudes of ulcer patients, at the risk of admitting some items that would be selected on the basis of chance alone. There were 10 additional significant items which dealt specifically with gastrointestinal disturbances or general physical health. Since the method of selecting the control group excluded both psychosomatic disorders and functional stomach complaints, these 10 items were not included.

For each attitude the writer has arbitrarily selected those significant items which seemed to be best subsumed under that attitude.[6] There is no claim that the choices of attitudes are necessarily the correct ones for the individual items. The only argument is that the items, when considered in their entirety as they are answered by the ulcer group, fit into the personality structure that is suggested by the totality of these attitudes.

---

[5] Seventy-four items represented 13.7 per cent of the 540 items (550 minus 10 somatic items). The difference of 3.7 per cent is 2.78 times the standard error of 10 per cent for 540 items. The difference is therefore significant at the .01 level of confidence.

[6] A minority of the items should properly be regarded as factual rather than attitudinal. However, even those items which pertain to the patient's overt behavior are valuable in evaluating his perception of himself.

The number preceding each item refers to the number of the item in the MMPI booklet form. The letter following the item indicates whether the item was answered more frequently "true" or more frequently "false" by the ulcer group than by the control group.

a. Psychological inadequacies are rigorously denied.

79. My feelings are not easily hurt. (T)

84. These days I find it hard not to give up hope of amounting to something. (F)

100. I have met problems so full of possibilities that I have been unable to make up my mind about them. (F)

152. Most nights I go to sleep without thoughts or ideas bothering me. (T)

171. It makes me uncomfortable to put on a stunt at a party even when the others are doing the same sort of things. (F)

172. I frequently have to fight against showing that I am bashful. (F)

264. I am entirely self-confident. (T)

297. I wish I were not bothered by thoughts about sex. (F)

299. I think that I feel more intensely than most people do. (F)

301. Life is a strain for me much of the time. (F)

310. My sex life is satisfactory. (T)

328. I find it hard to keep my mind on a task or job. (F)

338. I have certainly had more than my share of things to worry about. (F)

366. Even when I am with people I feel lonely much of the time. (F)

384. I feel unable to tell anyone all about myself. (F)

389. My plans have frequently seemed so full of difficulties that I have had to give them up. (F)

396. Often, even though everything is going fine for me, I feel that I don't care about anything. (F)

397. I have sometimes felt that difficulties were piling up so high that I could not overcome them. (F)

416. It bothers me to have someone watch me at work even though I know I can do it well. (F)

487. I feel like giving up quickly when things go wrong. (F)

521. In a group of people I would not be embarrassed to be called upon to start a discussion or give an opinion about something I know well. (T)

b. Femininity is denied and masculinity is exaggerated.

74. I have often wished I were a girl. (F)

203. If I were a reporter I would very much like to report news of the theater. (F)

223. I very much like hunting. (T)

239. I have been disappointed in love. (F)

240. I never worry about my looks. (T)

283. If I were a reporter I would very much like to report sporting news. (T)

295. I liked "Alice in Wonderland" by Lewis Carroll. (F)

423. I like or have liked fishing very much. (T)

435. Usually I would prefer to work with women. (F)

441. I like tall women. (F)

513. I think Lincoln was greater than Washington. (F)

563. I like adventure stories better than romantic stories. (T)

c. Masochistic, self-depreciating attitudes are denied.

79. My feelings are not easily hurt. (T)

262. It does not bother me that I am not better looking. (T)

459. I have one or more bad habits which are so strong that it is no use in fighting against them. (F)

d. Dependence is denied, but, at the same time, there is a disinclination toward self-assertion. One must be capable, but must not proclaim his individuality.

165. I like to know some important people because it makes me feel important. (F)

212. My people treat me more like a child than a grown-up. (F)

224. My parents have often objected to the kind of people I went around with. (F)

235. I have been quite independent and free from family rule. (T)

239. I have been disappointed in love. (F)

245. My parents and family find more fault with me than they should. (F)

264. I am entirely self confident. (T)

400. If given the chance I could do some things that would be of great benefit to the world. (F)

415. If given the chance I would make a good leader of people. (F)

441. I like tall women. (F)

e. There is an emphasis on proper and conventional behavior in order to avoid censure from others.

49. It would be better if almost all laws were thrown away. (F)

93. I think most people would lie to get ahead. (F)

99. I like to go to parties and other affairs where there is lots of loud fun. (F)

215. I have used alcohol excessively. (F)

410. I would certainly enjoy beating a crook at his own game. (F)

437. It is all right to get around the law if you don't actually break it. (F)

460. I have used alcohol moderately (or not at all). (T)

485. When a man is with a woman he is usually thinking about things related to sex. (F)

f. An impunitive rather than intropunitive or extrapunitive attitude toward aggression exists. However, aggression is expressed when such aggression would be socially approved or would lead to a feeling of moral self-righteousness.

28. When someone does me a wrong I feel I should pay him back if I can, just for the principle of the thing. (T)

59. I have often had to take orders from someone who did not know as much as I did. (F)

93. I think most people would lie to get ahead. (F)

137. I believe that my home life is as pleasant as that of most people I know. (T)

157. I feel that I have often been punished without cause. (F)

212. My people treat me more like a child than a grown-up. (F)

224. My parents have often objected to the kind of people I went around with. (F)

245. My parents and family find more fault with me than they should. (F)

253. I can be friendly with people who do things which I consider wrong. (T)

277. At times I have been so entertained by the cleverness of a crook that I have hoped he would get by with it. (F)

380. When someone says silly or ignorant things about something I know about, I try to set him right. (T)

383. People often disappoint me. (F)

426. I have at times had to be rough with people who were rude or annoying. (T)

438. There are certain people whom I dislike so much that I am inwardly pleased when they are catching it for something they have done. (F)

469. I have often found people jealous of my good ideas, just because they had not thought of them first. (F)

g. Moralistic and puritanical attitudes toward sex and religion prevail.

37. I have never been in trouble because of my sex behavior. (T)

58. Everything is turning out just like the prophets of the Bible said it would. (T)

98. I believe in the second coming of Christ. (T)

133. I have never indulged in any unusual sex practices. (T)

258. I believe there is a God. (T)

297. I wish I were not bothered by thoughts about-sex. (F)

378. I do not like to see women smoke. (T)

485. When a man is with a woman he is usually thinking about things related to sex. (F)

h. There is an extroverted, socially oriented "philosophy of life" in contrast to an introverted, introspective, ruminative attitude.

11. A person should try to understand his dreams and be guided by or take warning from them. (F)

152. Most nights I go to sleep without thoughts or ideas bothering me. (T)

207. I enjoy many different kinds of play and recreation. (T)

236. I brood a great deal. (F)

292. I am likely not to speak to people until they speak to me. (F)

366. Even when I am with people I feel lonely much of the time. (F)

384. I feel unable to tell anyone all about myself. (F)

Three items differentiated between the groups in a direction opposite to the predictions. They were:

132. I like collecting flowers or growing house plants.

452. I like to poke fun at people.

493. I prefer work which requires close attention to work which allows me to be careless.

The ulcer group answered the first two of these "false" less frequently than the control group did. The third item was answered "true" less frequently by the ulcer group. Despite the significance of these differences, the majority of the ulcer patients answered "false" to the first two of these items and "true" to the third one. Therefore, none of the 74 items was answered by the majority of the ulcer patients in a manner that was inconsistent with the predictions.

It is possible, of course, that some of the items which emerged as significant are chance items. The only rejoinder which can be made is that all but three of the differentiating items are consonant with the nature of the predictions. It would appear, therefore, that at least a large majority of the discriminating items cannot be attributed to chance. The objection may then be raised that the interpretations of the items are not unequivocal. Verbal statements are, after all, often ambiguous and may lend themselves to diametrically opposed clinical interpretations. To overcome this objection, the predictions were given to three advanced graduate students in clinical psychology who were told to accept these statements as accurate descriptions of ulcer patients' attitudes. They were then asked to answer the 74 items "true" or "false" as an ulcer patient would answer them. These judges were colleagues of the writer in a VA installation, but knew nothing of the purpose and nature of the study inasmuch as the study had been conducted in a different installation. The disagreements between the majority of the ulcer patients' responses to these items and the judges' responses numbered 15, 8, and 7. The agreement between each judge and the ulcer patients was statistically significant beyond the .001 level of confidence as measured by chi square. The tetrachoric intercorrelations among judges were .72, .78, and .97. Substitution of the median correlation of .78 into the Spearman-Brown prophecy formula yields a reliability coefficient of .91.

None of the ulcer patients, as compared to seven in the control group, had *Mf* scores exceeding 80. Several patients in the control group were, or had been, overt homosexuals, and aesthetic interests in these patients would not be unexpected. The *Mf* scale is primarily an interest rather than a clinical scale, and although the scale does not always diagnose homosexuality effectively, it is true that homosexuals, because of their interests, usually score higher on this scale. As far as the writer was able to determine from the historical material at his disposal, none of the ulcer group was, or had been, an overt homosexual. Assuming this to be an accurate finding, its significance is difficult to evaluate. The writer does not know what the expected chance frequency of overt homosexuality would be if one selected 30 cases at random from the patient population of the San Francisco Mental Hygiene Clinic. The absence of homosexuality in the ulcer group could very well be attributed to chance. However, the finding is consistent with the theoretical formulation, i.e., ulcer patients deny feminine trends in themselves. It is also consistent to assume that their moralistic sexual attitudes would condemn such culturally deviant behavior as homosexuality.

VI. DISCUSSION

The present study raises as many questions as it answers. Considerable evidence has been offered that ulcer patients who enter psychotherapy in a Veterans Administration mental hygiene clinic tend, on the whole, to be overtly passive, at least when they are compared with nonulcer patients in psychotherapy. However, the results are generalizable only to ulcer patients in psychotherapy from comparable socioeconomic backgrounds. One reason for presuming the present experimental sample to be overtly passive was its relatively low socioeconomic status. A sample of ulcer patients from a higher socioeconomic group should produce very different results. For example, one would hypothesize an extremely high level of aspiration for Alexander's sample (1, pp. 109-120) on the assumption that the Alexander group was derived from a higher socioeconomic level than the present sample. In view of Alexander's description of his ulcer patients, experimental verification of this hypothesis would certainly be expected. The Alexander group and the present group might not differ markedly on attitudes concerning the self, as measured by the MMPI, since both groups perceive themselves as active and efficient. However, opposite predictions would be made for the Alexander sample on all those measures which purport to measure overt passivity. The writer would argue that the familial background of these patients places greater emphasis on social mobility and tends to reward, rather than punish, that kind of active and assertive behavior which leads to success in our culture. This group, then, would be overtly as well as attitudinally counteractive. Limitations of time have prevented the procurement of such a group and so, for the present, the lack of data relegates any statements about these patients to the realm of untested, albeit tenable, hypotheses.

Another conspicuous omission in the present study is the absence of data on other psychosomatic groups. The differences between ulcer patients in therapy and nonpsychosomatic patients in therapy have been fairly well substantiated, but the possibility remains open that the differences are not so much attributable to the existence in the experimental group of ulcers, *qua* ulcers, as they are to the presence of a fairly general psychosomatic type.

Nevertheless, the results of this exploratory study are encouraging. The obtained differences between the groups provide some justification for regarding ulcer patients in psychotherapy as having psychiatric problems distinct from those of nonpsychosomatic patients in psychotherapy.

VII. SUMMARY AND CONCLUSIONS

The present study was designed to compare the personality structures of ulcer patients in psychotherapy and neurotic patients in psychotherapy.

A review of the literature pointed to two fairly specific ulcer personality types: an attitudinally counteractive, overtly passive type and an attitudinally counteractive, overtly active type. A rationale was presented for considering the ulcer patients in this study as attitudinally counteractive and overtly passive.

The ulcer group consisted of 34 male World War II veterans in psychotherapy with a medically confirmed diagnosis of a gastric or duodenal ulcer. The control group consisted of 38 nonpsychosomatic, nonpsychotic male World War II veterans who were also in psychotherapy. The groups were roughly equated for severity of psychopathology, length of time in psychotherapy, ethnic origin, intelligence, and socioeconomic status. The groups differed significantly with respect to age, but this variable did not correlate significantly with any of the experimental measures. The theoretical formulation of the ulcer personality led to specific hypotheses, which were tested by four measurements:

1. *Level-of-aspiration experiment.* The experimental task utilized performance on a pursuit rotor. Responses which represented a raising of one's goal after failure or a refusal to lower one's goal after failure were considered as atypical responses indicative of a high level of aspiration, and responses which involved a lowering of one's goal after success or a refusal to raise one's goal after success were considered as atypical responses indicative of a low level of aspiration.

The hypothesis was formulated that the ulcer group would have a lower level of aspiration than the control group when the level of aspiration was measured by the difference between the number of atypical high responses and the number of atypical low responses.

The results verified the hypothesis at the .05 level of confidence.

2. *Experiment on memory for completed versus interrupted tasks.* The experimental material consisted of 16 pencil-and-paper tasks. Half of the tasks were interrupted before completion.

The hypothesis was formulated that the number of completed tasks recalled relative to the number of incompleted tasks recalled would be greater for the ulcer group than for the control group.

The results verified the hypothesis at the .01 level of confidence.

3. *Experiment on perseveration of inadequate hypotheses.* Each patient was exposed to a learning problem with randomly distributed reinforcements for half of the first 24 trials, following which no more reinforcement trials were given. Two hypotheses were formulated:

a. The first hypothesis was that the ulcer group would require a greater number of trials to restructure the problem than the control group. The results verified the hypothesis at the .05 level of confidence.

b. The second hypothesis was that, after the problem had been restructured, the ulcer group would follow a smaller number of erroneous hypotheses before achieving the correct solution than the control group. The results did not verify the hypothesis.

4. *The Minnesota Multiphasic Personality Inventory.* On this test the hypotheses were formulated that the ulcer group would score higher on the Hysteria scale and lower on the Masculinity-Femininity and Psychopathic Deviate scales than the control group.

The results verified the hypotheses on the Hysteria and Masculinity-Femininity scales at the .01 level of confidence. The hypothesis that the ulcer group would score lower on the Psychopathic Deviate scale was not verified.

An item analysis confirmed the hypothesis that the ulcer patients in this study, despite their overt passivity, perceive themselves as active, masculine, and efficient, to an extent greater than in the control group. The item analysis also demonstrated that these patients are, on the average, conforming, moralistic, impunitive, and socially rather than introspectively oriented.

REFERENCES

1. Alexander, F., & French, T. M. *Studies in psychosomatic medicine.* New York: Harper, 1950.

2. Fenichel, O. *The psychoanalytic theory of neurosis.* New York: Norton, 1945.

3. Hathaway, S. R., & McKinley, G. C. Manual

for the Minnesota Multiphasic Personality Inventory. Minneapolis: Univer. of Minnesota Press, 1943.

4. HUMPHREYS, L. G. Effects of distribution of reinforcement upon the acquisition and extinction of conditioned responses. Unpublished doctor's dissertation, Stanford Univer., 1938.
5. MEEHL, P. E., & HATHAWAY, S. R. The *K* factor as a suppressor variable in the Minnesota Multiphasic Personality Inventory. *J. appl. Psychol.*, 1946, **30**, 525-564.
6. ROSENZWEIG, S. An experimental study of "repression" with special reference to need-persistive and ego-defensive reactions to frustration. *J. exp. Psychol.*, 1943, **32**, 64-74.
7. RUESCH, J., *et al. Duodenal ulcer*. Berkeley and Los Angeles: Univer. of California Press, 1948.
8. SCODEL, A. The psychodynamics of peptic ulcer patients; an experimental investigation. Unpublished doctor's dissertation, Univer. of California, 1951.
9. TORREY, J. Task completion as a function of organizational factors. *J. Pers.*, 1949, **18**, 192-205.
10. WECHSLER, D. *The measurement of adult intelligence*. Baltimore: Williams and Wilkins, 1944.
11. WOLF, S., & WOLFF, H. G. Evidence of the genesis of peptic ulcer in man. In S. S. Tomkins (Ed.) *Contemporary psychopathology*. Cambridge: Harvard Univer. Press, 1946. Pp. 163-174.
12. ZEIGARNIK, B. Über das Behalten erledigter und unerledigter Handlungen. *Psychol. Forsch.*, 1927, **9**, 1-85.

(Accepted for publication January 23, 1953)

Vol. 67, No. 11 Whole No. 361, 1953

Psychological Monographs: General and Applied

Democratic Leadership in the College Classroom[1]

Donald M. Johnson and Henry Clay Smith[2]
Michigan State College

I. INTRODUCTION

THE GENERAL purpose of the present study was to test the effectiveness of democratic leadership. This major problem of our society has been studied in the laboratory, in industrial and community groups, and in the classroom. Lewin and his students (10) conducted the pioneer laboratory experiment in this field; more recently Leavitt (9), Heise and Miller (6), and Mintz (12) have completed relevant laboratory experiments. Among the major industrial studies are those of Bavelas (2), Coch and French (3), and Maier (11). Studies of classroom leadership have been conducted and related studies reviewed by Asch (1) and by Smith and Dunbar (13).

The published research on this topic shows that the experimental arrangements are particularly difficult. Small groups can be formed and put to work under well-controlled laboratory conditions, but such groups are temporary and the group's activities are seldom of any genuine significance to the members. In on-going business and community groups, on the other hand, social processes may be observed under natural conditions but experimental variation of the important variables is seldom possible.

The college classroom offers a happy solution to some of these problems of experimental planning. The classroom group is not a temporary artificial one, but a regular part of the student's routine for one term. Social processes during class and the outcome of these processes are significant events in his life. At the same time, the social conditions of the classroom are to some extent under the experimenter's control, especially if the experimenter is also the teacher. The experimenter can control the membership of the class, its size, meeting place, and the like, and can exert considerable influence on the goals of the group, procedures used for attaining these goals, and other social activities during the class period. In addition, most of the group's social activities are open to observation, and data on the members' attitudes and achievements are easily obtained.

Furthermore, the college classroom deserves careful study and evaluation for its own sake, regardless of its contribution to social psychology. The classroom procedures of the present study are commonly used or recommended. Information on their relation to educational outcome should be of significance for educational practice as well as for the psychology of social interaction.

An adequate examination of the general belief in the superiority of democratic processes requires the statement of specific hypotheses in a form suitable for experimentation. It was expected that testing a series of related hypotheses would yield a general picture of the nature and effectiveness of democratic processes. The hypotheses tested were:

---

[1] Financed in part by the All-College Research Committee of Michigan State College.

[2] The order of names is alphabetical. Either author will be glad to supply details omitted for lack of space.

1. Students evaluate democratic classes more favorably than lecture classes.
2. Students accept decisions they make themselves more readily than decisions made by the instructor.
3. Students develop more democratic attitudes in the democratic class than in the lecture class.
4. Students learn more in a democratic class than in a lecture class.

II. METHODS

General Experimental Design

The design of the experiment is outlined in Table 1. In general, this design permitted a comparison between two classes taught by democratic methods and two classes taught by lecture-discussion methods. Variations between students were balanced by matching democratic and lecture groups. Variations between instructors were balanced by having each of two experimenters teach a democratic and a lecture group. Pretests and end tests were employed as well as various measurements during the course. The details of this design are discussed here in the order in which they appear in Table 1, subjects, tests, the experimental period, and the end tests.

Selecting and Matching of Subjects

In the winter quarter of 1951 at Michigan State College two classes in beginning psychology were scheduled and each allowed to build up to an enrollment of about 80. These two classes took a battery of attitude and achievement tests on the first three days of class. On the basis of the test scores a group of 16 students was matched with another group of 16 in the same class. One of

TABLE 1

General Experimental Design

		Democratic Groups		Lecture Groups	
		Instr. A	Instr. B	Instr. A	Instr. B
Subjects (grouped by pretest scores)					
Demo. Att.	*Achieve.*				
High	High	4	4	4	4
High	Low	4	4	4	4
Low	High	4	4	4	4
Low	Low	4	4	4	4
Total Number		16	16	16	16
Pretests (January)		60-item Democratic Attitude Scale and 150-item Achievement Test			
Experimental Period (10 weeks)		Use of democratic methods Measurements of participation, sociometric status, social passiveness		Use of lecture methods Measurements of social passiveness	
End Tests (March)		Democratic Attitude Scale, Achievement Test, Class Evaluation Scale, Test of Creativity			

these groups was then withdrawn from the class and met thereafter as a democratic class in a small conference room. The other group of 16 remained in the large class as a lecture group. In the other large class of 80, a democratic group and a matched lecture group were assembled in the same way. Thus instructor A taught, in a democratic way, an experimental group of 16 students while instructor B was teaching, by the lecture method, a large class of about 64 which included a matched control group of 16. The next hour, instructor B taught a democratic class of 16 while instructor A taught a large class which included the lecture group of 16.

The matching was done on the basis of scores from two tests, one of attitude toward democratic social processes and one of achievement in psychology, both of which will be described later. In addition to the matching of democratic and lecture groups, some subgroup matching was planned in order to permit comparison of the progress of those initially high on each test with those initially low.

At the close of the first three days of testing those students with high scores on both attitude and achievement were picked out and arranged into two groups of four each, roughly matched as to individuals and closely matched as to averages on both variables. Those with low scores on both tests were likewise arranged into two groups of four each. Then similar groups were made up of those who were high on attitude toward democratic social processes and low on achievement and those who were low on attitude and high on achievement. Thus we have in each group of 16 students four subgroups of four each. And each subgroup is equated to the corresponding subgroup in the other groups of 16. The groups as wholes will be designated Democratic Groups A and B and Lecture Groups A and B, according to their subsequent treatment and their instructor. About the middle of the term one student dropped out of instructor A's democratic group. His match was therefore withdrawn from the corresponding lecture subgroup in subsequent comparisons.

Starting the groups and subgroups off with approximately the same pretest scores had the advantages of equating opportunity to improve, reducing irrelevant variance, simplifying statistical procedures, and making posttest comparisons much easier to interpret.

Scales and Tests Employed

As indicated in Table 1, a Class Evaluation Scale, a Democratic Attitude Scale, and a 150-item achievement test with five subscores were used in both the pretest and end test periods. In addition, a test designed to measure one aspect of creativity, sensitivity to psychological problems, was employed in both democratic and lecture groups at the end of the course. During the course, records were also kept of individual participation, sociometric status, social aggressiveness, and attendance. Each of these measuring instruments is described here in the order just mentioned.

The Class Evaluation Scale. Students' evaluations of the democratic and lecture classes were obtained by the 20-item evaluation scale below.

CLASS EVALUATION SCALE

Directions. The following questions apply to your opinions about this class. They are requested as an aid in improving the course. Use the *separate* answer sheet. Mark the number opposite the question which most clearly corresponds to your attitude about the course. Please try to answer every question.

1. About how many members of the class do you know by their first name?

2. How pleasant have the classes been?

3. How interested are you in the general purpose of the course?

4. In your opinion, how interested is the class as a whole in the purpose of the course?

5. In your opinion, how effective has the class been in advancing its purpose?

6. How satisfied have you been with the class up to the present time?

7. Sometimes a particular course is very important to a student. The goodness or badness of the class is a very significant matter. At other times, a particular course may have little or no importance. How would you rate the significance of this course for you?

8. Sometimes students become so "involved" in a class that they rarely think of themselves at all. At other times, they become only mildly involved, thinking mainly of personal matters. How would you rate your "involvement" in the course?

9. Have you felt free to express your ideas in class?

10. How frequently have you actually expressed your ideas in class?

11. What is your general impression of the quality of teaching?

12. What would you consider was the class's over-all impression of the quality of the teaching?

Directions. Indicate your observations of the teacher of *this* class by marking the appropriate space on the answer sheet opposite the number of the statement. Use the following code for each of the statements: 1. Never. 2. Seldom. 3. Occasionally. 4. Frequently. 5. Always.

13. He made good plans for the class.

14. He made good decisions.

15. He could see both sides of a question.

16. He could get students to want to work.

17. He was very interested in the success of the class.

18. He explained why he did things.

19. He could make tactful criticism.

20. He treated each student as an individual.

Directions. Indicate whether you approved or disapproved of the following aspects of the course. Use the following code: 1. Strong disapproval. 2. Mild disapproval. 3. Indifferent or don't know. 4. Mild approval. 5. Strong approval.

21. The objective type of tests used.

22. The method of determining grades in the course.

23. The number of tests given.

24. The particular days on which quizzes were scheduled.

25. The classroom procedures employed.

We would appreciate your comments on anything that struck you as particularly good or bad about the course. Please write these on the back of the answer sheet.

Each of the above 25 statements had five alternative answers. For reasons of space the alternatives are not given for the first ten questions. The items in this scale were adapted from those of Hemphill (7). The first 10 questions are designed to measure group social dimensions defined by Hemphill. Questions 11-20 pertain to the leader's behavior. Hemphill found leader qualities, questions 13-20, most highly related to rated effectiveness of a group. In addition, questions 21-25 were added to determine attitude toward procedures which the students were given power to decide in the democratic groups. This information provided data for checking the second hypothesis. The students were also asked to write their personal impressions of the course in an informal way. *A general evaluation* score was obtained by adding items 1 through 20. A *decision evaluation* score was obtained by adding items 21 through 25.

The Democratic Attitude Scale. A major hypothesis of this experiment concerned the changes in the democratic attitudes of students exposed to a democratic atmosphere. An adequate test of this hypothesis requires a definition of democracy and a scale to measure changes in students' attitudes toward democracy. Unfortunately neither a usable definition nor a scale was available. In meeting this difficulty a definition of the term, based upon the usage of social scientists, was attempted. Although complete agreement with this definition is not expected, statement of a working definition will at least make clear the basis for development of the scale of attitude toward democratic leadership and democratic social procedures.

A survey of the relevant literature indicates that the term "democracy" is used on three levels. It is used to describe social institutions, such as a democratic government or, more generally, a democratic organization. It is used to describe the leadership roles played by individuals in dealing with a face-to-face group. Recently it has also been used to describe personality, as, for example, the "democratic personality" as contrasted with the "authoritarian personality." It is generally assumed that there is a close relationship between democracy at these levels. Democratic individuals become democratic leaders and create democratic institutions, and, vice versa, democratic institutions mold democratic leaders and democratic personalities. Because of this assumed intimate relationship between these levels, a brief definition at each level is attempted, although the primary concern of the present study was the second level, the role played by the democratic leader.

Institutional democracy is the sense in which the term is generally most familiar. The difficulties with definitions at this level have been clearly pointed out by Spitz (14). He concludes, however, that the essential elements of democratic organization are the following: (*a*) Power is invested in the people. (*b*) Majority decision must prevail. (*c*) Minorities shall remain free to oppose. (*d*) Leaders are controlled by the followers.

Psychologists have been primarily interested in democracy as an aspect of the leader's role in a face-to-face group. The following outline, which was used as a guide in conducting the democratic classes and developing the democratic attitude scale, is a synthesis of statements made by psychologists in a variety of contexts.

1. The basic principle of democratic leadership is that *the leader acts in a group-centered way*. He acts in accordance with the expressed needs of the group rather than with his conception of how things should be done. The original work by Lewin, Lippitt, and White (10) was based on this principle. In their democratic group the group determined policies and procedures with the guidance and encouragement of the leader.

2. The democratic leader *assists in the expression and clarification of the needs of the group*. He may do this by listening permissively to the expressions of individuals, by reflecting and interpreting their expressions, and by protecting minority opinions. He evaluates the *work* of persons, rather than persons as such.

3. The democratic leader attempts to *aid in the satisfaction of expressed needs*. He does this, in part, by clarifying the limits within which the group has control, by delegating his decision-making power to the group, by submitting facts relevant to decisions, and by suggesting alternative actions.

4. The democratic leader attempts to *decrease his leadership role* and increase his membership role. Although seldom explicitly stated, this "trend" aspect of democratic leadership is implied in most uses of the concept by psychologists.

The "democratic personality" is a topic of more recent interest to psychologists than democracy as an institution and as a leadership technique. Research to date suggests that a democratic personality is liberal, equalitarian, scientific, and secure.

A democratic attitude scale, based on the rationale just described, was devised by one of the authors to measure attitudes toward democratic leadership in face-to-face groups. In the development of this scale the role of the democratic leader as outlined in the paragraphs above was used as a guide. The intention was to make this scale a generalized one, covering all group situations rather than to confine it to the classroom situation. There were two preliminary forms of this scale, Forms A and B, and the final form, Form C, which is reproduced below. For disguise it was called simply an opinion questionnaire.

DEMOCRATIC ATTITUDE SCALE

Directions. The following statements refer to opinions about how groups should be managed. Some people agree and others disagree with them. In stating your own personal feelings, consider yourself in a concrete situation such as a club, a classroom, or a political or business group. Think of a group of about *twenty* to *twenty-five* people.

Please mark on the separate answer sheet according to your agreement or disagreement, as follows: 1. Strong opposition, disagreement. 2. Slight opposition, disagreement. 3. Uncertain. 4. Slight support, agreement. 5. Strong support, agreement.

1. What a member of a group thinks is not too important as long as he is doing his job well.

2. A group should be able to count on the ability of its leader to arrive independently of the group at decisions related to their actions.

3. One goal of the group leader should be to become stronger in his leadership role and more detached from his membership role.

4. It is usually unwise for a leader to ask members of his group for suggestions before setting up an important project.

5. It is best for a leader to tell inefficient workers to "get busy or get out."

6. An agenda set by the group should be kept open for later changes or additions.

7. Knowing a great deal about an individual's home life is a great help in selecting the right person for a responsible job.

8. Reliance upon one particular individual as a leader should decrease as the group becomes more mature.

9. In participating in a group, one should always seek the opinions of other members.

10. It is almost always desirable for a leader to explain the duties and responsibilities of his job to the people in his group.

11. The average member of a group takes little pride in what he does on his job.

12. Groups where the meeting is planned by the leader usually make more progress than those where it is planned together with the members.

13. Only carefully chosen people should be taught how to influence others, because of the chance that such skills may be used to promote personal goals.

14. When it is possible for a leader to use either an individual or a committee, he should use the individual in order to insure efficiency.

15. It is frequently necessary for a leader to change mature adult minds without their knowledge when he judges it to be for their final good.

16. It is generally wise for a leader to prohibit talking between workers on a routine job.

17. There is little value in asking members of most groups to judge their leader.

18. The most important requirement of a good leader is a complete understanding of the job he has to do.

19. A good way for a leader to prevent further trouble among members of a group is to make an example of one of them.

20. Compromise between group members with strongly opposed points of view generally weakens the resulting group decision.

21. Before each important action that the leader takes, he should obtain the agreement of the group involved.

22. The diverse opinions of every member of the group should be given careful consideration before a decision is reached.

23. It is desirable that a representative from each committee chosen by the members of a group get together with the leader in regular meetings.

24. A group should generally not go so far as to exclude a member who threatens to disrupt the functioning of the group.

25. Evaluation of the work of each member by the entire group is a very valuable contribution to the functioning of the group.

26. It is sometimes undesirable for a leader to talk over the methods of approach to a project with the members of the group.

27. It is usually better for a few people in a group to make decisions and then to present them to the rest of the group.

28. It is usually best for a leader not to explain in detail changes in rules and policies to the members concerned.

29. There are generally some people in a large group who should not be accepted on the same level of equality as others.

30. The real test of any leader's way of dealing with other people is how well he gets the job done.

31. It is more important for a leader to be sensitive to the needs of the group than to be an authority in the field of the group's interest.

32. It is not the leader's task to provide the group with all the necessary information for the group to arrive at a decision.

33. Every member of a group should be a potential leader.

34. A group member should feel free to contradict the leader, even when it means involving the group in a long argument.

35. The leader's primary task is to reflect the feeling of the group rather than to direct its thought and action.

36. A discussion-type group wastes far more time than one in which the leader makes most of the decisions.

37. In a group meeting it is a waste of time to have to listen to a group member who obviously knows less about the subject than the leader.

38. If a group is to run smoothly the leader must be dominant.

39. Group participation methods are fine

ideals, but for many purposes they are too inefficient to be useful.

40. A group leader in politics should be essentially a person capable of inspiring personal loyalty in the group members.

41. It is the leader's task to try to find out what the group wants rather than to decide what is best for the group.

42. It is primarily the responsibility of the group, rather than the leader, to see that the group arrives at a good decision.

43. In an ideal group the only difference between the leader and the members would be that they had slightly different jobs to do.

44. A leader should be free to discipline a group member whose behavior is detrimental to the group.

45. It would be quite possible to use group participation methods in running a business.

46. People are generally either leaders or followers, and it is the leaders who must be singled out for responsibility.

47. A group leader in business should be essentially a person who is able to get high work output from the members.

48. A group leader in education should be primarily an informative lecturer.

49. The more authority a leader has, the smoother the group will run.

50. Group discussions are usually monopolized by people who like to hear themselves talk.

51. The quality of a leader is dependent upon his ability to get people to do what he thinks is right.

52. Because of the relatively low educational level of the people he represents, a union leader should make the decisions for his group.

53. A person in a position requiring leadership must be morally and intellectually superior to the people in the group.

54. The nature of an instructor's position makes democratic methods very difficult in the classroom situation.

55. An individual who is responsible for a group's decisions should not have to be bothered with individual personalities.

56. A leader must be careful to see that every side of an issue is presented to the group.

57. A member of a group should feel free to argue with other members of the group even if he is all alone in his position.

58. Members of a group will get more out of discussing a problem than listening to the leader, even when the leader is an expert in the field.

59. A leader should be very sensitive to the feelings of others.

60. Ideally, religious groups should arrive at decisions on church matters by using group discussion methods.

This scale was administered to all subjects in the present study. The corrected split-half reliability was .85. A test-retest correlation was obtained by correlating initial and final scores separated by a two-month interval. This correlation was .47, indicating some degrees of stability of the scores, but not enough to interfere with measurement of shifts in attitude.

Some evidence exists to support the validity of the scale. The test has face validity in that the items have been selected according to a rational definition of a democratic attitude and empirically tested for internal consistency. Some of the items are valid in the sense that they discriminate successful from unsuccessful leaders, as on the How Supervise? test (4) from which some of the items were taken. The results of the preliminary form were significantly related to some personality variables, in the directions predicted from the above discussion of the democratic personality. It was an incidental purpose of the present study to gather additional data on the validity of this scale.

Social Passivity Scale. The social adjustment items from the Bell Adjustment Inventory were administered once in each of the classes to obtain a measure of the student's social aggressiveness. Low scores on this scale indicate a positive and constructive approach to social situations; high scores indicate social passivity. In order to make the title consistent with the scoring it is referred to as the Social Passivity Scale.

The Achievement Tests. An adequate test of the fourth hypothesis would require several different measures of achievement, because the experimental variables might have different effects on different kinds of achievement. Several writers have pointed out the inadequacy of the conventional measures of educational achievement, saying that democratic teaching produces improvement in certain intangible ways not ordinarily tested. The present experiment, therefore, included a wide variety of measures of achievement.

A large pool of test items based on the textbook was available. Difficulty values and reliability data for each item had been obtained from previous use with similar populations. From this pool the most reliable items were grouped into three subtests of 60 items each, intended to test, respectively, psychological vocabulary, facts, and reasoning. Examples of each type follow.

Vocabulary item

SET (1) preparation (2) reaction (3) pathway (4) choice (5) fixation

Factual item

Raising body temperature by electric current is a specific treatment for (1) psychosomatic illness (2) psychoneurosis (3) brain syphilis (4) manic-depressive psychosis (5) paranoid schizophrenia.

Reasoning item

Under which of these conditions will the sex drive in rats be stronger than hunger or thirst? (1) Give plenty of food and water. (2) Place animal in obstruction box when young. (3) Deprive animal of food, water, and sex for two days. (4) Deprive animal of food and water for one day. (5) Keep male and female together constantly.

Items of each type were assembled into two forms of 30 items each of equal average difficulty values. Thus two equivalent forms of each of three kinds of test pertaining to the textbook were available for initial and final testing. Exact equivalence was not necessary since all students took Form A first and Form B last.

The fourth type of achievement test was a case study test modified from one developed by Horrocks and Troyer (8). Sam Smith's behavior in school is described, then interpretive statements and recommendations are made, such as "Sam feels quite insecure in social situations." The students are asked if they agree with the statement, if the evidence presented is not sufficient, or if they disagree. The same test of 30 items was used initially and finally.

The fifth achievement test was a reading test. Paragraphs, tables, and diagrams, taken from the psychological literature, are printed, and questions are asked about this material. Two forms of 30 items each, roughly equated for difficulty, were used as initial and final tests.

Test data—all subjects. The experiments began with two large classes of 80 to 90 students. At the end of the term achievement test data were complete for 162 subjects. Table 2 shows the means, standard deviations, and reliability coefficients for Form A, given at the beginning of the course, and Form B, given at the end. Reliability coefficients are odd-even correlations, corrected for length by the Spearman-Brown formula. Tables 3 and 4 give the product-moment correlation coefficients between the part scores before and after the course. These three tables, summarizing the test results for all 162 students, present the important data necessary for interpreting the results of the comparisons of the

TABLE 2

TEST DATA FOR 162 SUBJECTS BEFORE AND AFTER A QUARTER TERM OF THE BEGINNING PSYCHOLOGY COURSE

Test	Initial Test, Form A			Final Test, Form B		
	M	*SD*	Reliability	M	*SD*	Reliability
Vocabulary	12.8	3.3	.46	22.6	3.9	.76
Facts	10.9	3.3	.51	18.8	4.2	.68
Reasoning	15.6	2.9	.33	21.0	3.5	.70
Case study	18.8	2.9	.39	19.8	3.0	.67
Reading	21.7	2.2	.34	20.4	3.1	.67
Total	79.8	9.6	.62	102.6	11.6	.86

TABLE 3

CORRELATIONS BETWEEN INITIAL SCORES
(Coefficients above the diagonal are corrected for attenuation.)

Test	Voc.	Facts	Reas.	C.S.	Read.
Vocabulary		.66	.64	.28	.38
Facts	.32		1.09	.37	.09
Reasoning	.25	.42		.22	.33
Case study	.12	.16	.08		.11
Reading	.15	.04	.11	.04	

small matched experimental and control groups.

From Table 2 it is apparent that scores on the initial tests include much chance variance. A chance score on each of the first three tests, for example, would be six. Average obtained scores are only about double the chance scores. On the case study test a chance score would be ten, and on the reading test a chance score would be about fourteen. For this reason the reliabilities of the part scores are rather low. When all five tests are combined to get a total score, the reliability of this total is reasonably satisfactory. The initial matching was based on this total score on all five tests.

TABLE 4

CORRELATIONS BETWEEN FINAL SCORES
(Coefficients above the diagonal are corrected for attenuation.)

Test	Voc.	Facts	Reas.	C.S.	Read.
Vocabulary		.67	.44	.06	.14
Facts	.48		.77	.18	.41
Reasoning	.32	.53		.28	.69
Case study	.04	.12	.19		.39
Reading	.10	.28	.47	.26	

On the final tests standard deviations and reliability coefficients are higher, as one would expect. The means, of course, are much higher, but it is important to note that they are not near the ceilings of the tests, which would be 30 on each subtest. The distribution of neither the initial scores nor the final scores is skewed.

Actually, the reliability with which this experiment is concerned is not the reliability of any one test but the reliability of the gain or improvement. Knowing the correlation between initial and final test scores, which was .34, and the reliability of each test, the reliability of the gain scores can be computed according to a formula developed by Thorndike (15, p. 614) as .61. This reliability is high enough to disclose large variations in individual scores or small variations in means of matched groups.

The improvement on the first three tests is about what one would expect in a quarter's course. As stated earlier, Forms A and B are closely equivalent in respect to vocabulary, facts, and reasoning. The improvement on the case study test is very small, even though Form B was the same test as Form A. And the improvement on the reading test is actually negative. The reading test has not been as carefully standardized as the first three tests, so the negative improvement may be the result of differences in difficulty between Form A and Form B. In any event the last two achievement tests do not contribute very much to total achievement for the whole sample.

From Tables 3 and 4 it is apparent that success on each test is related to success on the others; all correlations are positive. But many correlations are rather low, a fact that justifies the attempt to measure these different aspects of achievement separately. The reasoning test, as might be expected, has the highest average correlation with the others, both initially and finally.

Both initially and finally the case study test is the most independent. Evidently this test measures an aspect of achievement that is not tested by conventional tests of achievement, and, furthermore, a course in psychology does not add much to this aspect of achievement. Further study of this type of test might improve the criteria of achievement in psychology courses and might point up a remediable defect in teaching procedures.

In general it is clear that there is much to be learned by including several dependent variables in an experiment on teaching procedures. The course does not have the same effect on all variables; in fact it leaves some aspects of achievement practically untouched.

Ratings of sensitivity to psychological problems. One of the advantages claimed for the democratic method is that it encourages creative imagination. This aspect of achievement is not included in the battery of tests described above, all of which are of the multiple-choice type, suitable for machine scoring. One free-response type of test was added, therefore, following a suggestion by Guilford.

Guilford says that "a large part of the scientist's success depends upon his ability to ask questions, and, of course, to ask the right questions." He therefore suggests, "One might present the examinee with a short paragraph of expository material and instruct him to ask as many questions as he can that are suggested by the statements, with relatively liberal time allowed" (5, p. 452). Two paragraphs were used, with ample space for responses.

1. The President of the Sophomore Class appoints a committee of seven students to make plans for a class dance. When the committee meets, practically all the talking is done by two people.

List as many psychological *questions* as you can think of that are suggested by this statement.

2. Two married women of about thirty decide to learn to type, each practicing alone in her spare time. After a month of such practice one types fast and inaccurately, the other slowly and accurately.

List as many psychological *questions* as you can think of that are suggested by this statement.

All classes took this test the last week of the term.

Tests of this kind can easily be scored quantitatively by simply counting the questions listed. However, a qualitative measure of the merit of the questions requires the aid of expert judges. The judging procedure will be described in the section on results.

Measurement of group processes. Group processes in the classes were analyzed both for the additional light they might shed on the hypotheses and for the possibility of checking previous findings on the nature of such processes. In implementing this purpose, records of the number of individual participations were kept, similar to those reported by Smith and Dunbar (13). The index used was the number of voluntary participations divided by the number of days present in class.

Sociometric ratings were obtained from both democratic groups as an aspect of a group project where individuals were to be divided into subgroups to study psychology. The members of each class were asked to indicate as many individuals as they wished with whom they would like to work on this project. Attendance records were kept in all classes. In addition, the instructors kept a log of classroom activities and the students made written comments on such activities and on the course as a whole.

The Classroom Procedures

All four classes were the same in some superficial respects. All used the same textbook, all had the same initial and final tests, and all met four times a week for a one-quarter term. The democratic procedures were carried out during class time, hence all groups had approximately equal contact with instructors outside class.

The two large classes, within which were the two control groups of 16 each, were taught by the usual lecture-discussion procedure. Both experimenters had taught the beginning course often and endeavored to teach these classes by their customary methods.

The democratic classes were conducted according to the experimenters' definition of the role of the leader in the democratic group, as outlined in the section on the measurement of democratic attitudes. The major aspects of this role are recapitulated here.

1. In general the teacher acts in a group-centered way. He acts in accordance with the expressed needs of the group rather than with his conception of how things should be done. For example, the experimenters announced to their democratic classes that their average grade would be approximately the same as in the large lecture classes. (This was done for control purposes.) However, the instructors suggested to them that it might be valuable to deviate from traditional testing procedures and utilize essay tests, ratings by the instructor, ratings by fellow students, or some combination of these. It was quite apparent to the instructors that the students felt anxious in this unfamiliar situation and were seeking conventional landmarks. When the students in both democratic groups proposed the use of objective tests, the instructors merely aided them in working out the details. In this case, the instructors were attempting to fulfill two aspects of the democratic role: to clarify the limits within which the group had control and, within those limits, to operate on the basis of the expressed needs and decisions of the groups rather than on the basis of their conception of how things should be done. An objective indication of the attainment of this aspect of the democratic role is given by the fact that both democratic groups felt that the instructors treated them more as individuals than did the control lecture groups.

2. The teacher assists in the expression and clarification of needs by listening permissively to their expression, protecting minority opinion, and by evaluation of ideas rather than individuals. For example, when the topic of attention was introduced, the question of causes for poor concentration was raised. This led the students to recount personal problems in trying to concentrate in the dormitories, in the library, and at home. Criticisms of the educational system in this connection were advanced. In such discussions the leader attempted to listen permissively, to reflect feelings, and occasionally to interpret them, but not to restrict the discussion.

3. The teacher attempts to aid in the satisfaction of expressed needs by clarifying the limits within which the group has control, delegating his decision-making to the group, submitting facts relevant to decision, suggesting possible alternative actions, summarizing agreements in the group, and keeping discussion on the point.

4. The teacher aims to decrease his leadership role and to increase his membership role.

To conduct college classes according to these democratic principles is a difficult assignment. No one could say, in advance, whether it could be well done. But the experimenters were interested in the project. It was their principal occupation for a term, and they met almost daily to discuss operational problems.

Room arrangements were particularly fortunate. While the control classes met in a large lecture room with the usual instructor's desk on a slightly raised platform, the experimental classes met in a small conference room. Tables were pushed together, and the members of the group, including the instructor, sat in a ring around the tables.

Within the first few days the students were asked to discuss, then vote on, the number and kind of quizzes, order of topics, and the like. The decisions made by the experimental classes about quizzes were imposed by the experimenters on the control classes. Throughout the course they made original proposals and decisions as a group about classroom activities, field trips, topics to be discussed, and methods of study. When the topic of group organization and leadership arose, the experiment served as a basis for discussion.

Comparison of class activities. The instructors kept logs of the principal activities of all classes. Activities of the two large lecture-discussion classes were similar to each other and similar to most other classes taught in the same conventional way. Activities in the two small classes differed somewhat from each other and differed greatly from the conventional classes.

The most obvious difference between the two types of classes was the difference in the way they spent their time. A visitor to a lecture-discussion class would have observed the instructor lecturing about psychological topics and discussing these with the class practically all the time. Visiting a democratic class he would have observed the students doing much of the talking, making most of the decisions. He would have noted that a great variety of topics was discussed. The democratic classes, for example, spent many hours discussing aims of the course, ways of studying, kinds of examinations, grading systems, and the like. One democratic class, taught by instructor A, voted to have, in addition to the final examination, a midterm examination and four quizzes. The other democratic class, taught by instructor B, voted to have a midterm examination and two quizzes.

In the democratic classes the instructors made no attempt to cover the material of the course. As noted above, the teacher's role was to help the students satisfy their needs. Some students, especially at first, wanted to sit and listen to the instructor, but class participation increased as they became more familiar with each other and with the democratic procedure.

The principal difference between the two experimental classes apparently was a difference in attitude toward grades. Class B seemed to be interested in grades and worried that they might not do as well as the conventional classes. Class A, though they voted to have regular quizzes, seldom discussed grades or the other classes, and were quite willing to take time off for extracurricular activities.

In general it was clear from the experimenters' classroom observations that the experimental groups did make most of the decisions, that they did express some of their needs, that the class activities were organized around these needs, and that the teachers did increase their membership roles. And it seems probable that these democratic procedures were accepted more wholeheartedly by Class A than by Class B.

III. RESULTS

Student Evaluation of the Classes

The first hypothesis states that students will evaluate democratic classes more favorably than lecture classes. The data from the Class Evaluation Scale, which was administered at the end of the course, were used in testing this hypothesis.

General results. Table 5 summarizes the results of this comparison. The figures in the table refer to the general evaluation score which was obtained by adding items 1 through 20. The maximum score was 100. Confidence levels in this and succeeding tables come from a two-tail test of significance.

TABLE 5

DIFFERENCES IN STUDENT EVALUATIONS OF DEMOCRATIC AND LECTURE CLASSES

Instructor	Demo. Class	*N*	Lect. Class	*N*	Diff.	Conf. Level
Instructor A	75.9	15	65.7	15	10.1	.001
Instructor B	66.7	15	70.0	15	−3.3	n.s.
Total	71.3	30	67.9	30	3.4	n.s.

Clearly the 15 students in instructor A's democratic class evaluated their class more favorably than the 15 matched students in his lecture class. For instructor B the difference is reversed, but not significantly so. When the classes are combined, the difference favors the democratic classes, but not significantly so. In other words, the hypothesis was fully verified in the subexperiment under instructor A, but was not verified in the subexperiment under instructor B.

Differences between the democratic groups. These results were unexpected. Therefore the individual items of the Class Evaluation Scale were analyzed separately in order to disclose the most significant difference between the two democratic classes. These differences are shown in Table 6, arranged in order of statistical significance. Only those factors that were significant beyond the 5 per cent level have been included. It is clear that group A found its class pleasanter, more interesting, and more satisfying. There is no evidence that the instructor variables are any more important than the social dimension variables.

Acceptance of Group Decisions

The second hypothesis of this study was that students would accept decisions they made more fully than the same decisions when they were made for them by the instructor. Previous studies have uniformly supported this hypothesis. The data on this point come from questions 21 through 25 of the Class Evaluation Scale.

Table 7 shows that democratic group A was significantly more favorable to-

TABLE 6

SIGNIFICANT DIFFERENCES IN STUDENT EVALUATIONS BETWEEN THE TWO DEMOCRATIC GROUPS

Variable	Demo. Grp. A	Demo. Grp. B	Diff.	Critical Ratio
Social Dimension Variables				
4. Interest	3.80	2.87	.93	4.27
2. Pleasantness	4.47	3.27	1.20	3.28
6. Satisfaction	3.67	2.47	1.20	2.88
8. Involvement	3.13	2.47	.66	2.29
Instructor Variables				
16. Stimulation	3.53	2.60	.93	3.75
11. Teaching (own opinion)	3.93	3.27	.96	3.44
12. Teaching (estimate of class opinion)	3.80	3.00	.80	3.29
Total: Items 1–20	75.87	65.80	10.07	3.43

TABLE 7

STUDENT ACCEPTANCE IN DEMOCRATIC AND LECTURE CLASSES OF DECISIONS MADE IN THE DEMOCRATIC CLASSES

Instructor	Demo. Class	N	Lect. Class	N	Diff.	Conf. Level
Instructor A	22.7	15	19.7	15	3.0	.01
Instructor B	17.5	15	19.4	15	−1.9	n.s.
Total	20.0	30	19.5	30	0.5	n.s.

ward the decisions made than their matched control group, whereas democratic group B was less favorable. When the two classes are combined, the difference practically disappears.

TABLE 8

SIGNIFICANT DIFFERENCES IN DECISION ACCEPTANCE BETWEEN THE TWO DEMOCRATIC GROUPS

Variable	Demo. Grp. A	Demo. Grp. B	Diff.	Critical Ratio
Decision Acceptance Variables				
22. Method of grading	4.40	2.13	2.27	7.59
25. Classroom procedures	4.53	2.93	1.60	4.57
23. Number of tests	4.67	3.53	1.14	3.06
21. Type of test	4.67	3.87	.80	3.05
24. Days of tests	4.40	3.53	.87	2.97
Total: Items 21–25	21.25	16.00	5.25	3.74

Table 8 compares the two democratic groups in respect to acceptance of specific decisions. The largest difference is in respect to the method of grading, and these differences in acceptance of decisions are larger than the differences in social dimension variables and instructor variables shown in Table 6. These differences in the attitudes of the two democratic groups toward the decisions they had made were observed and discussed by the instructors early in the course. The most likely explanation is that both groups were involved and took personal responsibility for the decisions. In group A they thus became involved in decisions which were satisfying and which resulted in greater satisfaction. In group B the students felt involved in decisions which turned out to be unsatisfactory, and this resulted in greater dissatisfaction than if these decisions had been made by the leader.

Gains in Democratic Attitude

The third hypothesis which this experiment was designed to check was that the students in the democratic classes would become more democratic in their attitudes than those in the lecture classes. The measurement of democratic attitude was made at the beginning and end of the course by means of the Democratic Attitude Scale.

Analysis of the results showed that there were only small, inconsistent changes in attitude in both the democratic and lecture groups. Variability of scores and of score changes were large, indicating that something happened during the course, but over-all comparisons of class means showed that neither the democratic nor the lecture experiences had any consistent effect.

Two incidental findings are worth noting. The democratic groups became more homogeneous in their attitude toward democratic social processes. This suggests a development toward common *norms* rather than toward democratic gains. Second, the low democratic groups under instructor B made significantly greater democratic gains than the low group under instructor A. The results in terms of students' evaluation were the reverse: the low democratic group under instructor A had the lowest evaluation. Under instructor A the low group found the class pleasant and satisfying, but under instructor B unpleasant and satisfying. The implication is that the frustration of undemocratic individuals in a democratic situation, though they find it unpleasant, leads to greater progress toward democratic attitudes.

Achievement Gains

The fourth hypothesis of this study stated that the experimental democratic procedures are more effective in promoting academic achievement than the usual lecture-discussion procedures. The data bearing on this hypothesis appear in Table 9 in the form of mean gain scores. Both democratic groups improved more than their matched control groups, but neither of these gains nor the gain for the two classes combined is significantly different from zero. We can only conclude that the experimental democratic procedures did not produce academic superiority in any large amount.

TABLE 9
ACHIEVEMENT GAINS IN DEMOCRATIC AND LECTURE CLASSES

Variable	Demo. Class	*N*	Lect. Class	*N*	Diff.	Conf. Level
Instructor A	26.3	15	24.6	15	1.7	n.s.
Instructor B	23.7	15	22.4	15	1.3	n.s.
Total	24.9	30	23.5	30	1.4	n.s.

The lack of inferior achievement in the experimental classes should also be noted. Certainly the experimental classes spent much less time on the conventional subject matter of psychology than the large lecture classes. If anyone asserts that students allowed to make their own decisions about classroom procedures will fall behind their colleagues in conventional classes by, say, four score units, our data compel the rejection of this assertion at the 1 per cent level of significance.

It might be assumed that those who were initially favorable to democratic classroom procedures would profit more by the democratically run classes than those who were initially unfavorable. A special analysis of the data refutes this assumption.

Relative achievement on the five subtests. Although the democratic classes taught by the two instructors came out with about the same superiority over their matched lecture classes, the pattern of relative gains on the five subtests differs considerably. Instructor A's democratic class exceeded his lecture class in vocabulary by a difference significant at the 5 per cent level. In instructor B's classes this difference is reversed, though it is not significant. Instructor A's democratic group also achieved superiority on the reasoning test by a difference that is significant at the 10 per cent level (or the 5 per cent level if a one-tail *t* test is used). Likewise the relative gains on the case study and reading tests are reversed. The differences may be due to differences between the instructors or, more likely, to differences in the activities undertaken by the students when they make their own decisions.

There is some consistency in respect to relative gains. Both democratic classes achieved the same gains as their control classes in respect to facts. And both classes bested their control groups in respect to reasoning. For the combined groups, reasoning shows the highest relative superiority; reading is next. Though these differences for the combined classes are not significant, as far as they go they agree with the expectation that democratic procedures exert the most influence on the more intellectual achievements.

Sensitivity to psychological problems. The results of the test of creative thinking described earlier were in the form of questions about psychological problems. Each test paper of a student from a democratic class was paired with a test paper of his match from a lecture class. The 31 pairs of papers were randomized and given to four advanced graduate students of psychology with the following instructions.

> Read these by pairs and check the one of each pair that shows the most psychological curiosity and insight. Don't be influenced by vague generalities. Look for specific plausible hypotheses.

The four judges thus made 124 comparisons, of which 70 favored the democratic classes. This frequency does not quite reach the 10 per cent level of confidence (or the 5 per cent level if a one-tail test is used). Actually the comparisons of three of the judges did not differ from chance. The fourth judge, who had done previous research on complex judgments, gave 23 out of 31 judgments favoring the democratic subjects, a frequency that clearly exceeds the 5 per cent level.

The most reasonable conclusion is that the democratic procedures did increase sensitivity to psychological problems. Clear proof of this point would require more time from the subjects and more training of the judges.

Relationships Among Criteria

It is often assumed that the various criteria of effectiveness of a classroom procedure will be closely related. The validity of this assumption is of considerable importance, both practically and experimentally. The status of this assumption in the present study is shown in the product-moment correlations of Table 10. These are based on the 31 subjects of the two democratic groups and arranged in order of size.

TABLE 10

CORRELATIONS BETWEEN CRITERIA

($N=31$)

Variables	r
Class Evaluation ×Acceptance of Group Decisions	.43
Acceptance of Group Decisions ×Democratic Gains	.40
Democratic Gains ×Achievement Gains	.01
Acceptance of Group Decisions ×Achievement Gains	−.15
Class Evaluation ×Democratic Gains	−.16
Class Evaluation ×Achievement Gains	−.29

The significant correlations are between general evaluation of the course and acceptance of group decisions and between acceptance of group decisions and gain on the democratic attitude scale. Achievement gains are not positively related to any other criterion but may be negatively related to evaluation of the course and acceptance of group decisions. More detailed analyses agree with these correlation coefficients. We cannot assume that success on one criterion has more than a low relationship to success on any other.

Group Processes

The data presented thus far do not tell the whole story. The data on hand, including scores for social passivity and records of sociometric choices and class participation, were analyzed in several ways so as to disclose differences in motivation and perception of the classroom situation. One method of particular value was the comparison of extreme groups. The five students in each democratic class who made the highest evaluations of the class (as compared to their controls) were compared with the five lowest in each class. The same comparisons of the ten highest with the ten lowest were made for acceptance of group decisions, gains in democratic attitude, and gains in achievement. The significant findings may be summarized:

1. Those students in both democratic classes who accepted the group decisions most willingly made significantly lower gains in academic achievement. Apparently they were group oriented while those who made high achievement gains were grade oriented.

2. Favorable evaluation of the democratic classes was associated with high sociometric status. Those who were most fully accepted by the group liked the class the best.

3. Those who were initially high in achievement made significantly greater gains in democratic attitude. A reasonable explanation is that they were secure academically and thus could afford to try something new.

4. High achievement gains were associated with low initial democratic attitude, low sociometric status, low participation, and social passivity. The isolated individualist made the greatest academic improvement.

In general, the data suggest that the satisfactions derived from group work are different from the satisfactions derived from individual competition for grades in the typical class. Some students apparently preferred the traditional pattern and acted in the democratic class as they normally did in a lecture class. The individual grading system used in the democratic classes encouraged this approach.

IV. INTERPRETATION

The most interesting results of the experiment were, first, that the various criteria of effectiveness of the classroom procedures were not closely correlated and, second, that the two democratic classes carried out different classroom activities and yielded somewhat different sets of objective data. Table 11 summarizes the evidence in a general way.

TABLE 11

Summary of Evidence on All Hypotheses

Hypothesis	Experiment		
	A	B	A & B
Class evaluation	++	−	+
Acceptance of decisions	++	−	+
Democratic attitude gains	+	−	−
Achievement gains	+	+	+
Psychological sensitivity	++	+	+

++ Significant difference in favor of democratic group.
+ Insignificant difference in favor of democratic group.
− Insignificant difference in favor of lecture group.

The results from the contrasting classes of the two experimenters, or the two experiments, agree in respect to the more intellectual criteria. In the course in general psychology, democratic procedures did as well as, or better than, lecture-discussion procedures in promoting academic achievement and curiosity.

But the democratic procedures were accepted more wholeheartedly in instructor A's classes, apparently because instructor B's students got involved in an unsatisfactory grading system. When the instructors run the classes, they can run them all the same, but when students run them, variations in procedure are to be expected. This dissatisfaction with the grading system in the B democratic class was generalized or displaced to the evaluation of other items and the acceptance of other decisions. Nevertheless the democratic classes were on the whole somewhat more favorably received than the lecture classes.

An analysis of group processes in the classes of both instructors indicated a polarity of acceptance of the unfamiliar democratic procedures. Group-oriented students accepted group decisions and were accepted by their fellows. Other students did not get involved in group activities but retained their achievement orientation, probably because they entered the class with less respect for the democratic procedures. They were socially passive, did not participate in class discussion, and were not chosen by their classmates. It is probably this difference in the orientation that the students bring to the class, or develop in the first few days, that determines what they get out of it and this accounts for the lack of correlation between the group criteria and the achievement criteria. Based on their analysis and interpretation of these results, the authors have formulated the following revised hypotheses to guide their future research:

The most effective class is one in which:

1. The content is organized to facilitate team activities.

2. Some group incentive is employed in the grading system. In other words, all students are given some reward based on the over-all achievements of their group or subgroup.

3. The instructor maintains a warm, objective relationship with students, and delegates considerable authority to them.

The effectiveness of such classes is most evident in creative and problem-solving activities and among students with initially favorable attitudes toward such procedures, who typically participate in their classes.

V. SUMMARY AND CONCLUSIONS

The purpose of the present study was to test the hypotheses that classes under democratic leadership are more effective than lecture classes in: (*a*) developing favorable evaluations of the class, (*b*) promoting acceptance of group decisions, (*c*) developing democratic attitudes toward group processes, and (*d*) improving ordinary academic achievement.

The subjects included a total of 62 students in four sections of general psychology. Two of the sections were under democratic leadership, and two were taught by the conventional lecture-discussion method. Each of the experimenters taught a democratic and a lecture class. All four sections were initially equated by matching subjects on the basis of scores on an achievement test and on a scale for the measurement of democratic attitudes. These instruments were administered again at the end of the course along with student course evaluation and group decision evaluation scales. Decisions made by the democratic classes in regard to tests, grading, and class procedures were utilized by the instructors in the lecture classes.

Evaluations of the class and acceptance of group decisions were significantly more positive in one democratic class compared with its matched lecture group, than in the other. For the other democratic group, the results were negative, but not significantly so. The difference in the results for the two democratic groups appeared to be due primarily to the fact that the negative group became preoccupied and dissatisfied with the grading system which they themselves had set up. Individually, those students who were selected sociometrically as informal leaders were most favorable toward the democratic class. This sociometric selection was significantly related to the extent of overt participation in the class.

Democratic attitudes did not change significantly in any of the classes and there were no significant differences between any groups. Democratic attitudes were measured by a scale developed for this project, having a split-half reliability of .85. Those individuals making the greatest gains on this scale were significantly higher in initial achievement. The data suggest that democratic attitudes became more homogeneous in the experimental or democratic group than in the lecture group.

Both democratic groups made greater achievement gains than their matched lecture groups, but the differences were not statistically significant. For the combined groups the reasoning subtest showed a higher superiority in favor of the democratic groups than any of the other subtests. The individuals making the greatest gains in total achievement were less democratic in their attitudes, more passive socially, lower in participation, and less often chosen sociometrically. The pattern of differences strongly suggests that the high achievers rejected the democratic process and focused upon the traditional classroom objective of making a good grade for themselves.

Both democratic groups did better in a test of "sensitivity to psychological problems," and the results were of borderline significance. This written test, adapted for the project, was evaluated by four independent judges who carried out a blind analysis.

Based on their interpretation of the results, the authors suggest as hypotheses to guide further research that the most effective class is one in which (*a*) the content is organized to facilitate team activities, (*b*) some group reward is employed in the grading system, and (*c*) the instructor maintains a warm, objective relationship with the students and delegates considerable authority to the class. The effectiveness of such classes will be most evident in creative and problem-solving activities, and among those students with initially favorable attitudes toward such procedures.

REFERENCES

1. Asch, M. J. Nondirective teaching in psychology: an experimental study. *Psychol. Monogr.*, 1951, **65**, No. 4 (Whole No. 321).
2. Bavelas, A. Group decisions in setting production goals. In N. R. F. Maier (Ed.), *Psychology in industry*. New York: Houghton Mifflin, 1946. Pp. 264-266.
3. Coch, L., & French, J. R. P., Jr. Overcoming resistance to change. *Hum. Relat.*, 1948, **1**, 512-532.
4. File, Q. W., & Remmers, H. H. Studies in supervisory evaluation. *J. appl. Psychol.*, 1946, **30**, 421-425.
5. Guilford, J. P. Creativity. *Amer. Psychologist*, 1950, **5**, 444-454.
6. Heise, G. A., & Miller, G. A. Problem solving by small groups using various communication nets. *J. abnorm. soc. Psychol.*, 1951, **46**, 327-335.
7. Hemphill, J. K. Relations between the size of the group and the behavior of "superior" leaders. *J. soc. Psychol.*, 1950, **32**, 11-22.
8. Horrocks, J. E., & Troyer, M. E. Case studies of ability to use knowledge of human growth and development. *Educ. psychol. Measmt*, 1947, **7**, 23-36.
9. Leavitt, H. J. Some effects of certain communication patterns on group performance. *J. abnorm. soc. Psychol.*, 1951, **46**, 38-50.
10. Lewin, K., Lippitt, R., & White, R. K. Patterns of aggressive behavior in experimentally created "social climates." *J. soc. Psychol.*, 1939, **10**, 271-299.
11. Maier, N. R. F. Improving supervision through training. *Indus. Relat. Res. Ass.*, 1949, **3**, 27-42.
12. Mintz, A. Non-adaptive group behavior. *J. abnorm. soc. Psychol.*, 1951, **46**, 150-159.
13. Smith, H. C., & Dunbar, D. S. The personality and achievement of the classroom participant. *J. educ. Psychol.*, 1951, **42**, 65-84.
14. Spitz, D. *Patterns of anti-democratic thought.* New York: Macmillan, 1949.
15. Thorndike, R. L. Reliability. In E. F. Lindquist (Ed.), *Educational measurement.* Washington, D.C.: Amer. Council on Education, 1951. Pp. 560-620.

(Accepted for publication February 24, 1953)

Vol. 67, No. 12 Whole No. 362, 1953

Psychological Monographs: General and Applied

The Prediction of Effectiveness as a Factory Foreman[1]

Martin M. Bruce[2]
Dunlap and Associates, Inc., New York, N.Y.

I. INTRODUCTION

THE LITERATURE is replete with articles on the subject of supervisors in industry. The importance of their position has been cited by Mandell (18) and Lawshe (14), among others. However, there is a paucity of psychological research on foremen specifically (1, 10).

II. THE PROBLEM

The primary purpose of this study was to determine the extent and nature of importance of certain skills, abilities, and personality characteristics in relation to effectiveness as a factory foreman. The five specific problems were:

1. Is the rating which served as the criterion multiple or simple?
2. If the criterion is multiple, what are the optimal weights assignable to the various factors? If the criterion is simple, what part or parts or combination will serve as the best standard?
3. What is the nature of the interrelationships among certain personality characteristics, skills, and abilities, on the one hand, and measures of effectiveness as a factory foreman on the other?
4. How adequately can effectiveness as a factory foreman be predicted by means of the battery of tests chosen?
5. What are the optimal weights assignable to each of the predictors?

III. METHOD

The Sample

The subjects in this study were 107 foremen in a major tobacco firm in the United States. This group constituted a 100% sample of the personnel who bore the title of foreman at the time of testing. The men were located in four separate buildings in a Southern city. Their ages ranged from 20 to 63, with a mean of 39.1 and a sigma of 8.99. All were employed by the company for at least one year and each was in the job of foreman for a minimum of six months. The maximum length of service was 17 years and the mean for the group was 10.1 years. The foremen who constituted this sample were nonunionized.

---

[1] This monograph is based on a thesis submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy at New York University (5).

[2] The author is indebted to Professor John J. Sullivan, Chairman of the Doctoral Committee, who aided in the formulation of the problem and saw the research to its completion. To Miss Ida Moore I am grateful for aid in the computations. Mr. Corlin O. Beum gave many helpful suggestions in connection with statistical procedures. Finally, my deep appreciation to my wife, without whose personal sacrifices this work could not have been completed.

Each foreman supervises a group of approximately 15 workmen engaged in particular phases of the company's operations involving the preparation of tobacco leaf into pipe or smoking tobacco. The men working under the direction of the foreman operate stemming, shredding, moistening, flavoring, sealing, stamping, and packing machines. The foreman interprets written and verbal orders and determines the procedure of work for his group. He receives instructions from the plant superintendent, assistant plant superintendent, and chief engineer.

He assigns duties to men and women working under his direction and inspects their work for quality and quantity. He assigns workers to machines and personally trains them in machine operation when necessary. The foreman keeps records of output for his department which he submits to the plant superintendent at the end of the work day. He is responsible for maintaining harmony among his group of workers and is responsible for their morale and health on the job. He institutes and carries out plant accident prevention programs by teaching safety through demonstrations and by continual observation for lack of adherence to the rules.

The foreman helps train subordinates, usually titled assistant foremen, and aids them during emergencies. He performs related duties of a supervisory and administrative nature. He occasionally operates machinery in order to maintain level of production in the absence of workers.

The Criterion Variables

The criteria employed in this study are ratings by management. These ratings are completed annually by the plant superintendent and personnel director of the factory. The assistant plant superintendent also completed the ratings on the foremen, but only the over-all score on this rating form was available in the company records. Areas covered by the rating include: leadership, knowledge, dependability and judgment, initiative and creativeness. The scale ranges from 75% to 100%, with the following descriptive adjectives offered on the form: 100%, exceptional; 95%, above average; 90%, average; 85%, below average; 75%, fair.

"Leadership" is divided into two subdivisions entitled "Personality Traits" and "Organization." These contain ten and four items respectively. "Knowledge" contains three items; "Dependability and Judgment" seven items; and "Initiative and Creativeness" eight items.

Scores for each of the four areas as well as a total score were employed as criterion variables. The score for a particular area, e.g., Knowledge, is the sum of the individual items in the section. The total rating is the mean score for the four parts. These criterion measures, in the above form, were available from management. One other measure was attainable in the company records. This was in the form of a composite score of the total ratings of the plant superintendent, personnel manager, and assistant plant superintendent.

The Predictor Variables

The predictors included age, years with the firm, education in number of years, and the following tests:

The Personality Inventory. This test has been widely used since its publication by the author, Robert G. Bernreuter, in 1931. By 1946 there were some 135 published studies pertaining to this instrument. The inventory can be scored for six factors: neurotic tendency, self-sufficiency, introversion, dominance, self-confidence, and sociability. In this study two of the six scores were not used, self-sufficiency and introversion. Split-half reliabilities reported by the author are based on student populations. They are: neurotic tendency, .91 and .88; dominance, .89 and .88; self-confidence, .86; sociability, .78.

The test is published by Stanford University Press at Stanford, California.

Ess-Ay Inventory. This test offers an over-all score purportedly indicative of persuasive ability. It consists of 155 personality items and has an estimated odd-even reliability of .86. This inventory is published by the Personnel Institute, New York, N.Y.

Otis Self-Administering Test of Mental Ability. This test has been employed extensively in educational as well as industrial organizations for many years. The 75 items, according to the author, measure verbal intelligence. The test may be administered for either 20- or 30-minute time periods. The longer administration time was used in this study. A test-retest reliability of .92 is reported by the test's author in the manual.

Copies of the Otis may be obtained from the World Book Company, Yonkers, New York.

Vocabulary Inventory. This test is a 100-item, multiple-choice, untimed form. The task required of the examinee is to select from among four alternatives the one which most nearly means the same as the key word. The test has an estimated odd-even reliability of .91. It is published by the Personnel Institute, New York, N.Y.

Social Intelligence Test. This test is the special edition of the five-part test originally published by George Washington University. This short form, published by the Psychological Corporation, contains only two parts entitled "Judgment in Social Situations" and "Observation of Human Behavior." Neither the manual published by George Washington University nor the one distributed by the Psychological Corporation offers any reliability data. Howard R. Taylor, in his review in the *Third Mental Measurements Yearbook*, reports a test-retest reliability of .89 for the five-section form.

B-B-S Inventory. The content of this test is a duplicate of the General Clerical Test published by the Psychological Corporation, with the exception of the omissions of the sections on filing and location of numerical errors. A difference also exists in administration procedure. In this test directions for the seven parts are read by the examinee during the time allotted for the completion of the section. The seven parts are timed individually. These parts consist of items covering the following: comparing and checking clerical copy, computation, reading, spelling, vocabulary, arithmetical problems, and grammar. The test-retest reliability for over-all score is .94. It is published by the Personnel Institute, New York, N.Y.

Mechanical Comprehension Test, Form AA. This test is a 60-item untimed form which utilizes a pictorial multiple-choice format to measure comprehension of mechanical relationships. Form AA is the lowest level of difficulty of the various forms of the test. The author reports a coefficient of reliability of .84 obtained by the split-half method. The Mechanical Comprehension Test is published by the Psychological Corporation, New York, N.Y.

Test of Mechanical Ability. This test was developed by Jack Hazlehurst at Northwestern University in 1940 on the basis of a factorial study of paper-and-pencil tests of mechanical ability. The test is of the paper-and-pencil variety and has four sections which are administered with time limits. These four sections are: spatial perception, the ability to distinguish between lengths of lines; spatial visualization, the ability to determine the direction in which a bolt moves in a block; tool recognition, the ability to identify common tools by name; and blueprint reading, the ability to indicate the lines missing in blueprint drawings. Reliability computed by the odd-even method is .82. This test is published by the Personnel Institute, Inc., New York, N.Y.

The 39 criterion and predictor variables and their numerical designations as used in the study are:

Over-All Average Ratings

1. Most recent over-all rating including the evaluation by the plant superintendent, assistant plant superintendent, and personnel director.
2. Previous year's over-all rating including evaluations by plant superintendent, assistant plant superintendent, and personnel director.

Personnel Director's Ratings

3. Personnel director's rating for leadership.
4. Personnel director's rating for knowledge.
5. Personnel director's rating for dependability and judgment.
6. Personnel director's rating for initiative and creativeness.
7. Personnel director's over-all rating.

Plant Superintendent's Ratings

8. Plant superintendent's rating for leadership.
9. Plant superintendent's rating for knowledge.
10. Plant superintendent's rating for dependability and judgment.
11. Plant superintendent's rating for initiative and creativeness.
12. Plant superintendent's over-all rating.

Census Data

13. Age.
14. Years of education.
15. Years with the company.

The Personality Inventory

16. Neurotic tendency.
17. Dominance.
18. Self-confidence.
19. Sociability.

20. Ess-Ay Inventory.

21. Otis Self-Administering Test of Mental Ability.

22. Vocabulary Inventory.

Social Intelligence Test

23. Over-all score.
24. Judgment in social situations.
25. Observation of human behavior.

B-B-S Inventory

26. Over-all score.
27. Comparing and checking clerical copy.
28. Computation.

29. Reading.
30. Spelling.
31. Vocabulary.
32. Arithmetical problems.
33. Grammar.

34. Mechanical Comprehension Test, Form AA.

Test of Mechanical Ability

35. Over-all score.
36. Spatial perception.
37. Spatial visualization.
38. Tool recognition.
39. Blueprint reading.

IV. BRIEF SURVEY OF THE LITERATURE

Considering the quantity of literature that annually is published on clerical workers, various professions, executives, and other specific job categories, there is a dearth of psychological research with respect to the important group of foremen. Ghiselli and Brown (10), reviewing the literature, found 185 instances of reports of the effectiveness of intelligence tests in the selection of employees in various occupations, only nine of which dealt with the selection of supervisors. After surveying the literature covering experimental studies of supervisory selection, Boulanger (3) concluded that such investigations have been few in number.

Many have written on the subject of the use of tests and their value with the broad group of supervisory personnel (1, 2, 4, 5, 6, 7, 8, 9, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 31, 35).

One of the most recent studies of foremen reported in the literature is by Sparks (30). This investigation is particularly pertinent to the present study. Sparks reported on the value of the Bernreuter Personality Inventory in differentiating good from poor foremen in a large oil refinery in the South. He found that scores from self-confidence and sociability keys did not correlate significantly with a criterion derived from rankings of senior supervisors and they did not differentiate good from poor foremen.

Utilizing the same personality inventory, Hanawalt and Richardson (11) concluded that the chief difference between leaders and nonleaders is one of adjustment; the leaders are better adjusted in general than the nonleaders. The leaders are also more dominant.

V. PROCEDURE

The test data were collected over a period of three weeks. Groups of foremen were tested by a trained test administrator. The groups were formed on the basis of the individual's ability to be spared from his job for a full day. All eight tests were completed in a single day with standard rest sessions between forms and an hour for lunch. The criterion data were completed prior to the collection of test data. These data and personnel data were obtained from company personnel records.

The first step in the analysis of data consisted of recording all test scores, other variables, and criterion data. The first analyses were completed on the ratings which served as the criterion variables. To determine the value of these ratings as a measure of the effectiveness of foremen, and assuming that normal distribution of ratings is desirable, a test of goodness of fit to the normal curve of over-all ratings was made. A frequency distribution of the ratings was computed and cumulative percentages plotted on probability paper. A normal curve is indicated when the center portion of the resultant curve is a straight line. This is an empirical, not a statistical, determination of normality. As a check, the mean and standard deviation of the distribution of ratings were computed, these points were plotted on the same probability paper, and the line determined by these points was plotted. In a normal distribution, these two lines coincide.

The reliability of the criterion was determined by correlating the combined over-all ratings of the judges for the previous year with the combined over-all ratings of the same judges for the present period. This coefficient is to a degree a measure of test-retest reliability. However, perfect reliability cannot be expected even if the rating form were perfectly reliable, for the men must change somewhat in the course of a year. Interrater reliability coefficients were obtained by correlating the over-all ratings of the foremen prepared by the personnel director and plant superintendent. Additional measures of interrater reliability were obtained by correlating the scores of the plant superintendent for each of the four parts of the rating form with the corresponding scores of the other rater. This provided five measures of interrater reliability.

A final criterion from among the ratings available was chosen and this was employed in computing a Wherry-Doolittle multiple regression equation that would maximally predict the criterion selected. The multiple *R* was computed using the product-moment correlations whose computation is described below.

Pearsonian product-moment correlations were computed among all criterion variables and all predictor variables. There are 12 criterion variables and 27 predictor variables, a total of 39. All of the correlations were computed with the aid of "C-D Correlation Charts" which were devised by Cureton and Dunlap. While there are less time-consuming methods of computing *r*, this was selected because of the internal checks offered by correlation charts.

VI. RESULTS

Criterion Variables

A variety of statistical techniques was employed in order to determine the nature of the criterion measures. First it was desirable to know if the ratings prepared by the evaluators were normally distributed. The technique employed to determine this was the plotting of four cumulative percentages on normal probability paper:

1. Most recent over-all ratings by the personnel director, plant superintendent, and assistant plant superintendent combined.

2. Over-all ratings for the previous year by the personnel director, plant superintendent, and assistant plant superintendent combined.

3. Most recent over-all ratings by the personnel director.

4. Most recent over-all ratings by the plant superintendent.

From the analysis of the plots, as described in the Procedure, it may be concluded that the criterion measures are for practical purposes normally distributed.

The equivalent of test-retest reliability was determined by correlating the most recent over-all ratings with the previous year's over-all ratings. The resultant *r* of .842 seems sufficiently high to justify confidence. Measures of interrater reliability appear in the correlation matrix shown later. There are five reliability coefficients in this matrix. These are correlations between variables 7 and 12, 3 and 8, 4 and 9, 5 and 10, and 6 and 11. The first is a measure of rater agreement with respect to over-all rating. The correlation was .629, also sufficiently high to justify confidence. The other four correlations give us interrater reliability for factors of leadership, knowledge, dependability and judgment, and initiative and creativeness in that order. These correlations proved to be .539, .575, .593, and .642. It is evident that there is at least some consensus of opinion concerning the effectiveness of the foremen in the various respective areas.

A common limitation in multiple item ratings obtained in industry is that they are in effect not composed of multiple factors, but are ratings of a single factor. In order to determine if the various parts included in the over-all ratings were independent or all measures of the same thing, a crude cluster analysis was made. This consisted of computing the mean and standard deviation of the intercorrelations of the criterion variables. A low mean and a large sigma would suggest the operation of more than one factor. A high mean and a standard deviation small in comparison to the mean, a leptokurtic distribution, would suggest that all intercorrelations are equal to the mean value and the small fluctuations are probably due to chance or sampling errors. This would indicate that all variables are measuring the same thing, the one factor probably explainable in terms of halo.

TABLE 1

PER CENT CONTRIBUTION OF EACH PART OF RATING

Variable	Per Cent Contribution		
	Personnel Director	Plant Superintendent	Total
Leadership	9.86	12.73	22.59
Knowledge	12.74	14.78	27.52
Dependability and judgment	8.53	12.56	21.09
Initiative and creativeness	13.44	15.36	28.80
Total	44.57	55.43	100.00

An analysis was made in order to determine the percentage of contribution of each of the four parts of the ratings of the personnel director and plant superintendent. Table 1 shows the results of this analysis. It is readily noted that all parts contribute approximately the same amount to the total variance and thus there is no compelling reason to select from among them for the criterion to be used in the computation of the multiple regression equation.

Since we are dealing with variances, all of the parts could be added, thereby taking advantage of the nonoverlap of variances. Because the average of eight measures of the same thing is probably more reliable than any one, the single criterion selected should be the total. However, what is probably an even more reliable criterion than the total of these two men is the combined rating of the personnel director, plant superintendent, and assistant plant superintendent. This combined rating as an over-all score was available to the researcher and was utilized as the criterion to determine the value of each of the predictors. This criterion measure is variable 1.

Predictor Variables

Having determined the nature of the rating which served as the criterion, the next step was to determine the nature of the intercorrelations among the personality characteristics, skills, and abilities and the various parts of the ratings. Tables 2, 3, and 4 present the results of correlating the test scores, age, years of education, years with the company, and criterion measures.

If we reverse the correlations involving those factors that have a negative connotation, e.g., neurotic tendency, we find that practically all correlations are positive. And this is to be expected. It is a common research finding that ability measures tend to be positively correlated; an individual skilled in one area tends to be competent in others. The more intelligent individuals tend to have better clerical skills, larger vocabularies, better mechanical comprehension, etc. (32, 33, 34).

Each of the 27 predictor variables has face validity. The fact that these 27 vari-

TABLE 2

CORRELATIONS BETWEEN CRITERION AND PREDICTOR VARIABLES*

Variable	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39
1	−.200	.108	−.109	−.098	.200	−.073	.075	.238	.290	.109	.133	.193	.099	.255	.208	.230	.186	.103	.101	.230	.182	.214	.265	.154	.246	.179	.194
2	−.121	−.056	−.086	−.096	.156	−.110	−.205	.103	.199	.113	.034	.119	−.005	.181	.075	.130	.136	.121	.079	.201	.156	.158	.174	.095	.133	.157	.093
3	−.243	.215	.054	−.148	.242	−.144	.096	.234	.328	.230	.244	.318	.196	.279	.334	.231	.244	.242	.124	.218	.210	.265	.351	.188	.379	.126	.270
4	−.056	.112	.273	−.087	.153	−.083	.048	.191	.210	.168	.130	.169	.097	.230	.341	.214	.155	.241	.099	.268	.111	.230	.315	.270	.288	.162	.234
5	−.031	.071	−.106	−.100	.141	−.060	.135	.156	.219	.148	.135	.171	.123	.210	.236	.176	.213	.111	.114	.242	.141	.200	.213	.057	.219	.116	.005
6	.189	.135	−.034	−.010	.197	−.086	−.010	.246	.240	.132	.177	.242	.135	.240	.260	.183	.222	.184	.086	.229	.209	.365	.326	.218	.273	.239	.253
7	.157	.149	.078	−.116	.194	−.095	.065	.219	.258	.171	.180	.225	.144	.282	.312	.236	.223	.227	.114	.288	.195	.298	.329	.204	.309	.174	.265
8	−.245	.099	−.241	−.078	.171	.051	.034	.188	.250	.052	.074	.218	.009	.156	.094	.143	.111	.014	.027	.124	.191	.161	.130	.057	.115	.125	.048
9	−.166	.138	−.021	−.141	.213	.098	−.004	.237	.243	.107	.105	.175	.192	.214	.178	.162	.159	.087	.113	.160	.203	.142	.150	.096	.131	.441	.078
10	−.187	.129	−.287	−.086	.159	−.052	.071	.216	.183	.045	.054	.119	.014	.192	.024	.122	.183	.050	.065	.189	.180	.107	.081	.045	.089	.062	.019
11	−.195	.104	−.169	−.004	.129	.002	.031	.176	.306	.120	.157	.225	.337	.227	.154	.152	.217	−.008	.115	.205	.246	.265	.214	.072	.183	.210	.174
12	−.218	.092	−.196	−.073	.158	−.059	.040	.202	.253	.080	.098	.179	.050	.198	.118	.140	.184	.122	.078	.169	.218	.174	.152	.031	.143	.181	.074

* Correlation coefficients .191 and greater are significant at the 5% level. Correlation coefficients .224 and greater are significant at the 1% level.

ables and the 12 criterion measures are practically all positively correlated suggests that the criterion measures have validity.

An important consideration with respect to the correlation matrix is the faith that can be placed in each individual correlation. Fisher's *z* transformation was used, indicating that coefficients .191 and above are significant at the 5% level. Those that are .224 and greater are significant at the 1% level. Seven predictor variables are significant at the 1% level while 13 of the 27 predictors, almost half, are significant at the 5% level.

Age. In examining the first predictor variable, age, we find that it is negatively linearly correlated with the criterion. There is no indication of curvilinear relationship with the criterion. The correlation is −.200, significant at the 5% level. The size of this *r* may be in part due to the emphasis today in industry on youth. The younger worker is often looked upon favorably because of his youth.

In this population there is a tendency for the younger man to be better educated, which is to be expected in light of the rise in educational level through the years in this country. The younger men tend to score higher on the tests of knowledge, skill, and ability. Of the 20 variables covering these factors, nine are significant at the 5% level, suggesting that in this population age is of significance in performance on these tests.

The younger foreman's responses on the Personality Inventory suggest a more extroverted, dominant, and confident person. However, of these three personality measures, only the correlation between dominance and age reaches a level of significance that gives us some assurance that the *r* was not obtained by chance. The older men in this popula-

TABLE 3

INTERCORRELATIONS OF CRITERION VARIABLES*

Variable	1	2	3	4	5	6	7	8	9	10	11	12
1		.842	.764	.681	.744	.809	.834	.820	.824	.806	.825	.877
2			.601	.561	.606	.746	.706	.785	.812	.724	.834	.826
3				.738	.253	.791	.287	.539	.522	.516	.500	.554
4					.659	.722	.883	.421	.575	.382	.407	.466
5						.569	.851	.522	.578	.593	.526	.593
6							.893	.581	.653	.564	.642	.639
7								.580	.616	.579	.577	.629
8									.744	.860	.847	.933
9										.711	.825	.873
10											.778	.904
11												.936

* Correlation coefficients .191 and greater are significant at the 5% level. Correlation coefficients .224 and greater are significant at the 1% level.

tion tend to be less persuasive and tend not to do as well as the younger foremen on the timed tests. There is a positive correlation between score on the vocabulary test, which is untimed, and the age factor. Clinical findings indicate that vocabulary tends to hold with age while skills of various sorts tend to diminish. This trend appears to be the same in this population, but it should be noted that the *r* does not reach the 5% level of significance. Nevertheless, the trend is definitely suggestive, since this correlation is positive while the others are negative.

Years of education. Years of education is a variable that is correlated positively with all test variables if we again reverse the B1N and F1C scores on the Bernreuter so that they have positive connotations. Years of education correlated highest with scores on the mental ability test, which might well be expected for this population as well as for any other. Its correlation with the criterion is not sufficiently high, only .108, to expect that it could not have arisen by chance. The moderate positive correlation may be due to management's tendency to view the number of years of education of an employee as a favorable factor.

Years of education is significantly highly correlated, 1% level, with all of the nonpersonality tests, except for Part III score on the Mechanical Ability Test. On the Bernreuter Personality Inventory the B4D score correlates with years of education to the extent that the *r* is significant well beyond the 1% level. It may be inferred then, at least for this population and these tests, that education is an important factor in success on tests of ability, skill, and knowledge. However, it may be of little or no importance in doing well as a factory foreman.

Years with the company. The number of years the foreman has been with the organization is of course closely associated with his age, as shown by the correlation of .526. For the most part, we find that the relationships between age and other predictor variables tend to be similar to the relationships between years with the firm and these same variables. However, there are some modifications.

The older men who have been with the firm for some time tend to do better in the mechanical areas than those older men who have not been with the firm for as long a period of time. There appears to be some relationship between the rating received from members of

TABLE 4

INTERCORRELATIONS OF PREDICTOR VARIABLES*

Variable	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39
13	−.412	.526	.125	−.278	.162	.017	−.245	−.332	.079	−.110	−.333	−.017	−.201	−.139	−.296	−.212	−.089	.039	−.049	−.280	−.269	−.338	−.182	−.369	−.089	−.291
14		−.294	.083	.356	−.146	.321	.267	.667	.453	.483	.453	.576	.478	.307	.453	.459	.327	.375	.242	.559	.318	.356	.246	.384	.078	.329
15			.097	−.145	.137	−.034	−.199	−.184	−.048	−.137	−.218	−.083	−.106	.151	−.163	−.131	.155	−.008	−.055	−.224	−.042	.038	.092	−.018	.223	−.061
16				−.752	.874	.285	−.269	−.163	−.211	−.070	−.200	−.014	−.093	.019	−.065	−.186	−.091	−.072	.120	−.163	−.034	−.031	−.002	−.127	.139	−.092
17					−.804	.006	.435	.267	.238	.144	.243	.280	.140	.058	.154	.211	.155	.071	−.005	.211	.072	.143	.162	.169	−.062	.194
18						.188	−.240	−.175	−.285	−.060	−.192	−.001	−.172	−.098	−.101	−.181	−.181	−.124	.052	−.219	−.028	−.056	−.014	−.146	.171	−.152
19							.042	.136	.137	.224	.023	.219	.064	.004	.141	.122	.058	.100	.095	.104	−.042	.060	.044	.035	.053	.036
20								.184	.047	.157	.220	.108	.166	.132	.184	.185	.125	−.065	−.006	.191	.151	.299	.154	.084	−.071	.115
21									.620	.680	.618	.626	.697	.383	.572	.601	.450	.603	.299	.666	.528	.535	.290	.517	.243	.550
22										.684	.627	.632	.331	.217	.168	.298	.298	.645	.138	.538	.325	.316	.250	.372	.036	.255
23											.731	.975	.614	.312	.436	.576	.343	.593	.261	.530	.360	.327	.187	.309	.138	.326
24												.568	.629	.376	.454	.482	.369	.519	.271	.483	.425	.458	.199	.421	.241	.479
25													.573	.271	.385	.546	.299	.539	.218	.487	.300	.260	.174	.247	.091	.237
26														.577	.712	.647	.760	.800	.643	.824	.436	.477	.269	.441	.317	.423
27															.407	.317	.337	.261	.299	.347	.321	.466	.246	.358	.405	.420
28																.393	.497	.461	.346	.591	.433	.463	.293	.448	.247	.378
29																	.287	.498	.260	.610	.518	.327	.107	.375	.167	.274
30																		.628	.393	.548	.166	.300	.258	.256	.133	.267
31																			.352	.646	.283	.269	.106	.260	.167	.264
32																				.466	.237	.238	.195	.183	.215	.188
33																					.369	.373	.194	.394	.201	.292
34																						.648	.374	.573	.457	.570
35																							.710	.874	.700	.732
36																								.464	.475	.376
37																									.390	.568
38																										.386
39																										

---

* Correlation coefficients .191 and greater are significant at the 5% level. Correlation coefficients .224 and greater are significant at the 1% level.

management and the foreman's length of service; those who are newer in the organization tend to be rated somewhat higher, but the correlation between this variable and the criterion is only —.109, which falls far short of the 5% level of significance. It should also be noted that the relationship between years with the firm and other variables tends to be modest. Only the *r* with grammar is significant at the 1% level. There are three correlations involving this variable that are significant at the 5% level in addition to the one mentioned above. One of these three is positive. The longer the foreman is with the company, the more tools he seems to recognize. The men with longer tenure seem to have less knowledge of grammar, less persuasive ability, and poorer judgment in social situations.

Neurotic tendency. Those in this population who show greater neurotic tendency as measured by the B1N key on the Bernreuter tend to be less proficient in mental ability, vocabulary, and persuasive ability. The correlations with virtually all variables are negative, including the *r* with the criterion. The correlation with the criterion, —.098, might well have arisen by chance.

Though the trend is in the direction of poor test scores for those with greater neurotic tendency, only one of the correlations is significant at the 1% level and two additional at the 5% level. The *r*'s with vocabulary and Part I of the Social Intelligence Test reach the 5% level of significance while the Ess-Ay Inventory score is significant at the 1% level.

As for the correlations with other Bernreuter factors, the findings for this population closely resemble those found in a group of 157 engineering students as reported in the manual for the inventory by Bernreuter. The correlations between B1N scores and B4D, F1C, and F2S scores in the engineering student population were respectively —.80, .95 and .32. In this population of factory foremen the correlations proved to be respectively —.752, .874, and .285. In the foreman group as well as in Bernreuter's student population, those showing greater neurotic tendency tended to be less aggressive, less confident, and more solitary or independent.

Dominance. The correlations between dominance as measured by the B4D key of the Bernreuter and variables purporting to measure skills and abilities and knowledge in various areas are positive with the exception of two. Both these negative correlations are close to zero and do not reach an acceptable level of significance. Of the 20 correlations between B4D and tests of a nonpersonality nature, five are significant at the 1% level and three more are significant at the 5% level. At the 1% level are the variables of persuasive ability, mental ability, vocabulary, judgment in social situations, and understanding of human behavior.

It appears that the more dominant foremen are those who are more capable in most of the areas measured by the tests employed. The importance of dominance in effectiveness as a factory foreman is shown in the correlation of .200 with the criterion. The *r* is significant at the 5% level but not at the 1% level. It appears to be a factor of some importance here.

Self-confidence. The F1C score is highly correlated with the B1N score in this population as it was in Bernreuter's engineering student population. He reported a correlation coefficient between these two variables of .95, whereas in this factory population the *r* is .874. Because of high correlation between these two variables, the correlations between other variables and F1C are very similar to the correlation coefficients between B1N and these same variables. In fact, the signs of the correlations are the same in all cases but one and the order of magnitude of the correlation coefficients is very similar. In the one instance of reversal of sign of *r*, both correlations are close to zero and might well have come about by chance.

In the case of the B1N variable, the *r*'s with three tests were significant at the 5% level and beyond. The same three scores are significantly highly correlated with F1C, Social Intelligence Test Part I, vocabulary and persuasive ability. The *r* between score for grammar and F1C reaches the 5% level of significance though it falls short of this level in the case of the B1N variable. Even the correlations with the criterion are similar, neither being sufficiently high to justify confidence in the correlation's not having come about through chance.

Sociability. The Bernreuter F2S factor is comparatively unrelated to the other Personality Inventory scores in this population. This was also true in the population reported on by the test author. Correlations between F2S and B1N, B4D and F1C in Bernreuter's student population were .32, .07, and .11 respectively. In this group of foremen the correlations in the same order are .285, .006, and .188.

The sociability factor does not appear to be of significant importance in success in the job of factory foreman and this variable appears to be unrelated to any significant extent with any of the other predictor variables except years of education and Part II and over-all score on the Social Intelligence Test. The *r* with education is .321, significant beyond the 1% level, suggesting that in this population the men with more education tend to be less sociable or more solitary and independent. The high-scoring individuals, those who tend to describe themselves as nonsocial, tend to demonstrate better observation of human behavior as measured by the Social Intelligence Test.

Ess-Ay Inventory. Score on the Ess-Ay Inventory is the first predictor variable to be correlated very significantly, at the 1% level, with the chosen criterion. It can be expected to contribute to the multiple regression equation, and does, as will be shown later. Because of the similarity in items, it can be expected that there would be some relationship between scores on this inventory and those obtained on the Personality Inventory. This is the case for three of the four scores obtained on the Bernreuter.

Correlations between persuasive ability and F1C, B4D, and B1N are significant at the 1% level. Those who score higher on the Ess-Ay Inventory tend to be less neurotically inclined, more dominant, and more confident. The relatively high linear relationship between dominance score and persuasive ability suggests that the Ess-Ay Inventory will tend to be related to other variables much the same as the B4D score. This is the case; the correlations are approximately of the same order of magnitude, and in all cases but one the signs are the same.

Mental ability. Mental ability as measured by the Otis Self-Administering Test of Mental Ability is the variable that is most highly correlated with the criterion. The *r* is .290; and of course it makes the greatest contribution to the multiple *R*. Mental ability in this population is correlated positively with all factors of ability, skill, and knowledge. In fact, the Pearsonian *r*'s between this test score and all other test scores of a nonpersonality nature are the highest in the matrix. No other test score correlates as highly with other test scores as does the raw score on the test of mental ability. Those foremen who do well on this test tend to do well on all others, whether timed or untimed. In this population, performance on the Otis appears to cover a considerable amount of what is measured by the other tests.

Vocabulary. It is a common research finding that knowledge of words correlates positively with measures of intellectual ability. The finding is similar within this group of 107 factory foremen. The correlation between scores on the 100-item vocabulary test and the Otis is .620, one of the highest in the matrix. Therefore, it is to be expected that vocabulary scores would correlate positively and highly with other test scores, as does the score on the Otis. However, the correlation coefficients are lower than the *r's* between Otis and the other test scores in all cases except four.

The 100-item vocabulary test, variable 22, correlates .645 with the 35-item timed vocabulary test which is part of the B-B-S Inventory, while the correlation between Otis scores and the 35-item vocabulary test, variable 31, is .603. Three of the four cases in which the vocabulary score is higher in its correlation with other test scores involve the Social Intelligence Test. However, the differences are modest at best, being .004, .009, and .006. If a generalization may be made from this population, it seems that the measures of "Judgment in Social Situations" and "Observation of Human Behavior" are to a large degree measures of verbal intelligence and verbal knowledge.

Vocabulary is positively correlated with the criterion but not to the extent that confidence can be placed in the *r*'s not having arisen by chance. Because of this and because vocabulary is so highly correlated with mental ability, which is most highly correlated with the criterion, it cannot be expected to be a valuable contributor to the prediction of effectiveness for this group in a battery of tests that includes the Otis.

Social Intelligence Test. Because the various scores obtained from the special edition of the Social Intelligence Test are highly correlated with scores on the Otis, it too cannot be expected to make an important contribution to the prediction of effectiveness as a factory foreman.

Part I of the Social Intelligence Test is highly correlated with Part II in this population, the *r* being .568. Part I is also highly correlated with the over-all score, the *r* being .731. However, it is Part II that contributes even greater variance to the over-all score. The correlation between Part II and over-all score on the Social Intelligence Test in this factory group is .975, the highest *r* in the correlation matrix. The *r* between Part I and the criterion reaches the 5% level of significance, but neither of the other two scores correlates sufficiently highly with the criterion for us to place confidence in the suggested positive relationship.

The over-all score on the Social Intelligence Test correlates positively with all other test scores of a nonpersonality nature, helping to demonstrate either that the foremen who are competent in one area tend to be competent in others, or that the tests to some extent are actually measuring the same thing, or a combination of these conclusions.

B-B-S Inventory. The pattern of positive correlations between test scores and the criterion and among the test scores is evident also in the relationships between the total clerical test score and criterion and between this test score and the other predictors. This is also true of the part scores on the B-B-S Inventory. The over-all score correlates sufficiently highly with the criterion to reach the 1% level of significance, suggesting that this total score or one or more of the parts may have value in predicting the criterion.

The average correlation between part score and total score on the B-B-S Inventory is .709, which is spuriously high because each part score is included in the total. However, it appears that in this population a clerical test consisting of fewer sections would probably do an adequate job. The average correlation with intelligence test score is .511. Part VI, consisting of arithmetical problems, variable 32, correlates lowest with the Otis score and next to the lowest with the total B-B-S score. Because of its comparative "independence" and because this variable correlates to a significant extent with the criterion, it proves valuable in helping to predict that measure of effectiveness.

Mechanical Comprehension Test, Form AA. The next test in the battery is Bennett's Mechanical Comprehension Test, Form AA. From the standpoint of face validity it would be expected that performance on this test would correlate higher with the criterion than most other tests in the battery. Actually, the correlation between Mechanical Comprehension Test scores and the most recent over-all rating chosen as the criterion is .033 points above the mean correlation between the 24 test predictor variables and the same criterion; the average correlation between test scores and the criterion being .181. Further, it should be noted that the correlation between test scores on the Bennett and the previous over-all rating is .032 points greater than the average correlation with the criterion, which is .126. These differences are small but tend to be in the direction of higher than average correlations with the criterion.

However, while the correlation between this test score and the criterion is significant at the 5% level, the test does not seem to have anywhere near the predictive value it apparently has in other situations. In the manual for the test, Bennett reports a correlation of .64 between scores made by machine tool operator trainees and a criterion similar to the one used in this study. Other research studies cited in the manual reveal correlation coefficients in factory populations of .52, .30, .41, and .55. Another limitation in the value of this instrument in the population of factory foremen included in this study is the fact that the correlation between mental ability score and performance on the Bennett Mechanical Comprehension Test is quite high, the Pearsonian r being .528.

Bennett reports a much lower correlation between the Otis Self-Administering Test of Mental Ability and his own test with high school students as the population. He reports an r of .25. However, the r is higher, .45, in a population of introductory engineering subjects course enrollees. Bennett does not further describe these populations and it may be that the 107 factory foremen represent a less attenuated group with respect to intellectual spread than the two reported by Bennett. This would explain the higher correlation with Otis in the group of foremen in this study.

Test of Mechanical Ability. The Test of Mechanical Ability (Hazelhurst's) on a face validity basis would also be expected to correlate more highly with the criterion than the other test scores. This is the case to a moderate extent for the over-all score on this test, as it is also with the Bennett Mechanical Comprehension Test.

The total score on the Test of Mechanical Ability correlates .265 with the most recent rating selected as the criterion and .174 with the previous rating. These are

.084 and .048 points higher than the respective mean correlations between other test scores and criterion measures. The over-all score on the Test of Mechanical Ability correlates highly with Bennett's Mechanical Comprehension Test, the *r* being .648. The relationship between the Mechanical Ability Test and the intelligence test score is also high, suggesting that this test measures to a considerable extent functions similar to those measured by the Otis.

All the part scores on the Test of Mechanical Ability correlate highly with the over-all score, but again this is a spurious correlation in that the part score is included in the total. The part scores are highly correlated among themselves, all to a degree greater than chance expectancy, suggesting that these are not actually independent factors in the factor-analysis meaning of the term.

The over-all score and the score for Part II of the Test of Mechanical Ability correlate with the criterion beyond the 1% level. Therefore, they might be expected to contribute to the multiple *R*, but do not because of the comparatively high order of correlations between these variables and the Otis scores, years of education, and age, three variables that make a contribution to the prediction of the criterion.

Wherry-Doolittle test selection method. The final step in the statistical analysis of the data was the application of the Wherry-Doolittle test selection method for the computation of the multiple *R*. This is a method for the selection of a battery of predictors that will give the maximum shrunken multiple correlation with the criterion after a correction has been made for the chance error added by each test. The tests and other predictors are selected in the order of their contribution to the multiple.

As a rule, the increase in the multiple becomes less and less as additional variables are included in the predictor team. Finally, the point is reached where the addition of another test adds more chance error than actual validity to the battery. Application of the Wherry shrinkage formula after the addition of each test will show when this point has been reached and no further additions are useful.

This method will not hold if two predictor variables have some part in common. When this is the case, spurious correlations are introduced. In the instances of variables 23, 26, and 35 we have over-all scores in which two or more parts are included. Therefore, in computing *R* by the Wherry-Doolittle test selection method, either these three variables or the 13 part scores must be omitted. The retention of the maximum number of predictor variables seemed desirable. Therefore, the three over-all scores were not included in the computation of *R*. The results of this computation indicate that the predictor variable that contributes most to the prediction of the criterion is the Otis score. The correlation of .290 between Otis score and the criterion is significant at the 1% level. This variable was the first to be selected in the Wherry-Doolittle process.

The prediction of the criterion is increased by .043 when the Ess-Ay Inventory score is added. We see then that of the factors measured, intelligence apparently is the most important in success as a foreman. Secondly, there is a personality factor that here is labeled "persuasive ability."

The third predictor that was found to contribute to *R* is variable 32, the Part VI (Arithmetical Problems) score on the clerical test. This section of the B-B-S Inventory contains arithmetical problems. With the inclusion of this variable R is raised from .333 to .360.

Again .027 is added to the multiple regression correlation coefficient when years of education, as a predictor, is added. Age is the next contributor to the extent of .011, raising the multiple R to .398.

The standard error of the multiple R is .088 when it includes only the Otis. When all five variables are included the standard error is reduced to .081.

The formula for the prediction of the criterion in terms of standard scores was:

Criterion = .484 (Otis) + .172 (Ess-Ay Inventory) + .167 (Arithmetical Problems) + .574 (education in years) — .143 (age).

If only the first test, the Otis, is utilized the standard error of estimate is 2.192. By adding the next four predictors the standard error of estimate is reduced to 2.102, a gain of but .09. As a result, though the multiple R becomes greater with the inclusion of additional variables there is actually little gain because the standard error is not appreciably reduced.

Assuming validity for the criterion measure, verbal intelligence is important primarily in predicting success and secondarily in predicting persuasive ability. This is reasonable on a face-validity basis. However, one would expect that some technical aspect of factory work would also enter the picture. The multiple regression equation does not bear out this expectation, possibly because: (*a*) technical knowledge and technical ability as measured by the mechanical tests employed are actually not important, or (*b*) we have not really measured technical knowledge and ability with the two mechanical tests, or (*c*) these technical qualities are actually to a large degree mental ability.

Beyond the Otis score, variables included in this battery of tests do not materially improve the prediction of the criterion. The lack of reduction in the standard error of estimate, in addition to other factors, suggests that management might well confine itself to a small number of variables rather than administering a comprehensive battery of tests when considering a man for the position of foreman.

If the selection ratio is high, i.e., if a few foremen are to be selected out of a large number of applicants, the weights may be used to advantage. If the selection ratio is small, there will be a limited improvement over chance selection because of the size of the standard error.

VII. SUMMARY

This study was designed to investigate the importance of certain personality characteristics, skills, and abilities in effectiveness as a factory foreman. It was the purpose of this study to determine the value of certain psychometric instruments and other predictor variables in predicting success in this job in the tobacco company involved in this study.

The first step was to determine the nature of the rating form used as the criterion, then to determine its value as such a yardstick. Its reliability could be determined in certain respects, but its validity could only be assumed because of the difficulty of assessing validity of such a measure. The reliability of the criterion measures proved sufficiently high to justify confidence in them. However, a crude cluster analysis of the parts of the rating form suggested that separate factors were not measured by the rating; the ratings appear to reflect considerable "halo effect."

Following this analysis, the linear relationships between each of the predictors and each of the criterion measures were determined. The nature of the interrelationships was expressed in terms of product-moment correlations. Using these data and the finally selected single criterion measure, the Wherry-Doolittle test selection method was utilized in order to maximally predict the rating, which consisted of the combined judgments of three qualified raters.

It was found that five predictor variables contributed to the multiple *R*. The following sequence gives the order of their selection by the Wherry-Doolittle method, together with the multiple *R* as each variable is added: Otis score (.290), Ess-Ay Inventory score (.333), Arithmetical Problems score (.360), years of education (.387), and age (.398). Age was negatively correlated with the criterion. A formula for the prediction of the criterion based on these five predictors was worked out on the basis of standard scores.

Of the 27 predictor variables, 13 were correlated significantly with the criterion, at the 5% level or better. Seven of these correlations reached the 1% level of significance. These seven variables in the order of their correlation with the chosen criterion are:

1. Otis Self-Administering Test of Mental Ability.
2. Test of Mechanical Ability, over-all score.
3. B-B-S Inventory, over-all score.
4. Test of Mechanical Ability, Part II, Spatial Visualization.
5. Ess-Ay Inventory.
6. B-B-S Inventory, Part VI, Arithmetical Problems.
7. B-B-S Inventory, Part II, Computation.

Many of the tests that were more highly correlated with the criterion did not enter the multiple regression equation, primarily because of a high correlation with intelligence, the first variable included. The majority of variables included in the analysis did not appear to be of significant importance in the job performed by these factory foremen. And, as in similar research studies, there were few variables uncovered that contributed to the prediction of a criterion such as the one selected.

It seems justified to conclude that in a situation such as the one described in this study, a large battery of tests cannot be justified even if the time is available. In this case the major contribution of .290 to the multiple *R* of .398 was made by a test purporting to measure verbal intelligence. Comparatively little of reliable value was added by the other tests.

Further research in this area should seek to find instruments that measure factors that can successfully predict the criterion despite their lack of correlation with intelligence.

REFERENCES

1. Achard, F. H., & Clarke, F. H. You can measure the probability of success of a supervisor. *Personnel J.*, 1945, **21**, 353-373.
2. Beckman, R. O., & Levine, M. Selecting executives: an evaluation of three tests. *Personnel J.*, 1930, **8**, 415-420.
3. Boulanger, J. Selection of supervisors. Unpublished master's thesis, University of Minnesota, 1949.
4. Brown, R. R. The Rorschach in industry. *Personnel*, 1948, **24**, 434-436.
5. Bruce, M. M. The importance of certain personality characteristics, skills and abilities in effectiveness as a factory foreman. Unpublished doctor's dissertation, New York University, 1952.
6. Clay, H. Experiences in testing foremen. *Personnel*, 1952, **28**, 466-470.
7. File, Q. W. Are management's views of supervisors faulty? *Personnel J.*, 1946, **24**, 242-250.
8. File, Q. W., & Remmers, H. H. The measurement of supervisory quality in industry. *J. appl. Psychol.*, 1945, **29**, 323-337.
9. File, Q. W., & Remmers, H. H. Studies in supervisory evaluation. *J. appl. Psychol.*, 1946, **30**, 421-425.
10. Ghiselli, E. E., & Brown, C. W. The effectiveness of intelligence tests in the selection of workers. *J. appl. Psychol.*, 1948, **32**, 575-580.
11. Hanawalt, N. G., & Richardson, H. M. Leadership as related to the Bernreuter Personality Measures: IV. An item analysis of responses of adult leaders and non-leaders. *J. appl. Psychol.*, 1944, **28**, 397-411.
12. Harrell, W. Testing cotton mill supervisors. *J. appl. Psychol.*, 1940, **24**, 31-35.
13. Kurtz, A. K. A research test of the Rorschach Test. *Personnel Psychol.*, 1948, **1**, 41-52.
14. Lawshe, C. H. How can we pick better supervisors? *Personnel Psychol.*, 1949, **2**, 69-73.
15. Maier, N. R. F. *Psychology in industry*. New York: Houghton Mifflin Company, Inc., 1946.
16. Mandell, M. M. Testing for administrative and supervisory positions. *Educ. psychol. Measmt*, 1945, **5**, 217-228.
17. Mandell, M. M. The group oral performance test. *Publ. Personnel Rev.*, 1946, **7**, 209-212.
18. Mandell, M. M. The selection of foremen. *Educ. psychol. Measmt*, 1947, **7**, 385-398.
19. Mandell, M. M. Selection of blue and white collar supervisors. *Personnel*, 1948, **24**, 321-327.
20. Mandell, M. M. Selection of supervisors. *Mgmt Rev.*, 1948, **37**, 79-80.
21. Mandell, M. M. Hypotheses regarding selection methods for supervisory and executive selection. *Amer. Psychologist*, 1950, **5**, 280-281. (Abstract)
22. Mandell, M. M. Improving your supervisory selection methods. *Mgmt Rev.*, 1950, **39**, 68-70.
23. Mandell, M. M. Research findings in the field of supervisory and executive selection. *Personnel J.*, 1950, **27**, 215-216.
24. Millard, K. A. A personnel study of supervisors in business and industry. Unpublished doctor's dissertation, University of Minnesota, 1947.
25. Moore, H. Supervisory selection. *Personnel J.*, 1942, **20**, 353-356.
26. Pederson, Ruth. Development of a paper-and-pencil test of supervisory judgment. *Amer. Psychologist*, 1949, **4**, 277-278. (Abstract)
27. Riegel, J. W. Selection and development of prospective foremen. Bureau of Industrial Relations, University of Michigan, 1941, *Bulletin No. 11*.
28. Sartain, A. Q. Relation between scores on certain standard tests and supervisory success in an aircraft factory. *J. appl. Psychol.*, 1946, **30**, 323-332.
29. Sinaiko, M. W. The Rosenzweig P-F Study in selection of department store section managers. *J. appl. Psychol.*, 1949, **33**, 36-42.
30. Sparks, C. P. Limitations of the Bernreuter Personality Inventory in selection of supervisors. *J. appl. Psychol.*, 1952, **35**, 403-406.
31. Stockford, L. Selection of supervisory personnel. *Personnel*, 1947, **24**, 186-199.
32. Super, D. E. *Appraising vocational fitness.* New York: Harper, 1949.
33. Terman, L. M., *et al. Genetic studies of genius: the promise of youth*. Palo Alto: Stanford Univer. Press, 1930.
34. Terman, L. M., & Oden, M. H. *The gifted child grows up*. Palo Alto: Stanford Univer. Press, 1948.
35. Thompson, C. E. Selecting executives by psychological tests. *Educ. psychol. Measmt*, 1947, **7**, 773-778.

(Accepted for publication March 3, 1953)

Vol. 67, No. 13 Whole No. 363, 1953

Psychological Monographs: General and Applied

Responses to Successive Auditory Stimuli at the Cochlea and at the Auditory Cortex[1]

Mark R. Rosenzweig[2] and Walter A. Rosenblith[3]

Psycho-Acoustic Laboratory, Harvard University

The sounds we hear are usually members of a series and are not isolated events. This sequential occurrence of stimuli is also characteristic of other sensory modalities, and a number of studies have been devoted to the perceptual effects of stimuli delivered in close succession. In audition, perceptual studies employing successive stimuli have dealt with such phenomena as short-time masking (34), auditory localization (40), the relative loudness of successive clicks (10), the precedence effect, and the perceptual fusion of successive sounds (44). As yet the neural mechanisms of these perceptual phenomena are not well understood, although several recent papers have provided information about the responses of the auditory cortex to successive acoustic clicks (7, 11, 28, 42).

We have undertaken the experiments reported in this monograph in order to explore further the neural processes involved in the perception of successive sounds. To do this, we have recorded electrophysiological responses in anesthetized cats at the two extreme stations of the ascending auditory nervous system—the round window of the cochlea and the auditory cortex. In the following sections we will describe and compare the events that occur at these two levels when successive click stimuli are presented. Following an initial click, the responsiveness of the auditory system to a second click is depressed. At the cochlear level, the recovery of responsiveness is monotonic; it can be described quite precisely in terms of a mathematical, albeit a statistical, model (32). At the cortex, description of the recovery of responsiveness seems to require at least two factors, one monotonic and the other cyclical, and there are further complications that prevent us from arriving at a simple formal description of the cortical events.

METHOD

Experimental Procedures

Electrophysiological responses were recorded in anesthetized cats at the round window of the cochlea or in its vicinity and at auditory areas I and II of the cortex. The amplified responses were photographed from a cathode-ray oscilloscope.

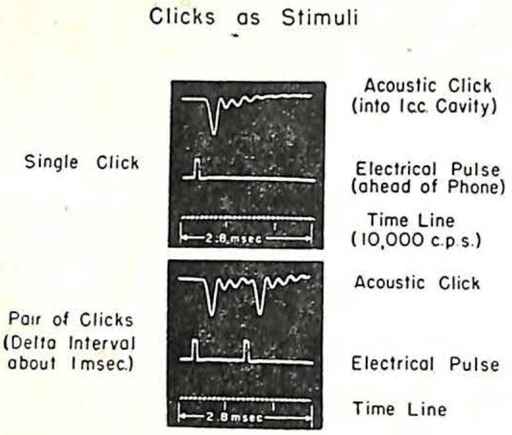
Stimulus equipment. (a) Click generator: produces electrical square pulses (Fig. 1) which have 0.1-msec. duration and 0.62-volt amplitude. Two clicks can be separated by any interval from 0 to 450 msec. The generator was built by Paul Dippolito of the Psycho-Acoustic Laboratory. (b) Earphones: two PDR-10 (Permoflux), one for each ear, matched by physical and psychological tests for pure-tone and transient responses, and for polarity. (c) Coupling from earphone to ear: flexible plastic tube, 0.4 cm. inner di-

---

[1] This research was carried out under Contract N5ori-76 between Harvard University and the Office of Naval Research, U. S. Navy (Project NR 142-201, Report PNR-117). Reproduction for any purpose of the U. S. Government is permitted.

[2] Now at the Department of Psychology, University of California, Berkeley.

[3] Now at Massachusetts Institute of Technology, Cambridge, Massachusetts.



11/29/1950

FIG. 1. Electrical pulses and acoustic clicks. An electrical pulse of 0.1-msec. duration is transduced into an acoustic click by an earphone. The click is conveyed by a plastic tube into the cat's ear. The output of the earphone-tube system, working into a cavity whose acoustic properties are presumably equivalent to those of a cat's ear, is pictured above. Note that for a delta interval of 1 msec. there is no appreciable interaction between the two acoustic clicks.

ameter, 6 cm. long, conducting sound from special housing on earphone into incision in external meatus near eardrum. Cotton yarn (3 cm. long) was placed in tube to reduce resonant peaks in the frequency response. (*d*) Calibration: acoustic signal delivered to 1-cc. cavity which represents an impedance reasonably equivalent to that of the ear of the cat. Signal recorded by Western Electric 640 AA condenser microphone. The frequency response of our stimulus system was flat within ±5 db from 50 to 4,500 cps. It is difficult to specify the click stimulus more precisely than by giving the dimensions of the electrical square pulse and the frequency response of the transducing system. Figure 1 shows that the acoustic click consisted primarily of a large brief pulse, followed by a train of small damped sine waves. In order to control the intensity of the click, the electrical square pulse could be attenuated in steps of 1 db. The output of the pulse generator (0.62 volt across the earphone) is 0-db attenuation, our click reference level. The intensity of clicks weaker than the standard will be given in terms of their attenuation: a stimulus of -60 db is one that is 60 db weaker than our standard. The threshold of human observers for clicks delivered through this acoustic system lies around 70 db below the click reference level.

Recording equipment. (*a*) Electrodes: platinum or silver wire. The indifferent electrode was attached to the head holder which gripped the animal's mouth. (*b*) Amplifiers: until the summer of 1950 we used a broad-band grid-to-ground amplifier of about 74-db gain built by Rufus Grason of the Psycho-Acoustic Laboratory. Thereafter we used two matched broad-band, low noise level differential amplifiers of about 72-db gain, built by Grason. Actually most of the round-window responses were recorded with a pass band of 100 to 5,000 cps, and most of the cortical responses with a pass band of 10 to 1,000 cps. (*c*) Cathode-ray oscilloscope: until the summer of 1950, a Dumont 247 CRO; thereafter, a four-channel CRO built by Grason and Dippolito. (*d*) General Radio oscilloscope camera, model 651-AE; speed control modified for stable operation over wide range of speeds by Ralph Gerbrands and Paul Dippolito.

Physiological preparation: Cats, usually adult. Anesthetic, usually Dial in Urethane, occasionally Nembutal.

The experiments extended from the summer of 1948 through the spring of 1951.

Acoustic Clicks as Stimuli

Neural potentials in response to click stimuli were first described by Davis and his group in the course of their investigation of the Wever and Bray effect. Further studies of the responses to clicks have been made at the cochlea (33, 34), at the brain stem (1, 30, 41), at the thalamus (18), and at the cortex (7, 11, 22, 23, 28, 35, 37, 38, 42). There are several reasons for using a click stimulus when one wants to investigate the electrical behavior of the auditory nervous system as seen through a gross electrode; that is an electrode that has certain averaging and integrating properties. A click resembles those brief pulselike disturbances that are the natural signals of the nervous system. Besides being natural sensory stimuli, clicks have almost all the advantages that go with direct electrical stimulation; a click is a sharp transient that evokes a recognizable response at all locations of the auditory system. Other acoustic stimuli, such as pure tones or thermal noise, do not yield responses that can be as directly and easily identified in the higher centers. In part this ready identification of click responses is due to their clear-cut temporal characteristics; the latency and duration of the responses increase in regular fashion as we ascend the auditory system. Moreover, responses to clicks are remarkably similar in such animals as cats, monkeys, guinea pigs, hamsters, bats, and pigeons. Finally, we can control the number and intensity of the clicks delivered and the time intervals separating them.

The advantages of click stimuli outweigh the difficulties of measuring their physical characteristics. There is little doubt that in the future we will have to control our stimuli more carefully; we may want to filter our clicks or even try the short tone pips other workers have used. But at present we feel that we can still learn a good deal from responses to common-purpose clicks, that is, responses to brief pulses that are equivalent to reasonably wide-band acoustic disturbances and that evoke almost simultaneous responses from an appreciable number of neural units.

The Combination of Round Window and Auditory Cortex

The work reported here includes data from the two extreme stations of the ascending auditory nervous system—the round window and the auditory cortex. These two stations have certain obvious advantages in terms of technique: electrodes can easily be located at them without interfering too much with the normal existence of the anesthetized cat. The choice of these two locations does not imply, however, that events important for the understanding of cortical responses do not occur at the intermediate brainstem or thalamic levels. Work at the two extreme stations emphasizes the contrast in the characteristic behavior of peripheral and central stations of the nervous system. It should not, however, be forgotten that this contrast is seriously contaminated, as is shown in this paper, by the much greater sensitivity of cortical responses to such parameters as level of anesthesia and temperature—parameters whose quantification at the present time remains rudimentary.

The strategic importance of our two locations is enhanced by the following considerations. Just as the microphonic components of the round-window response presumably afford a reasonably reliable monitoring device for the mechanical events in the inner ear, so the responses from the cortex seem to be as close as we can now come on the afferent side to observing electrical events antecedent to the behavior of the whole animal.

Last, but not least, the choice of locations is the outcome of the concern illustrated in the current literature with the neural components of the round-window response and with the responses of the auditory cortex. We have learned how to manipulate reversibly the so-called first neural component of the round-window response by means of temperature, masking, or fatigue (25, 29, 34), and how to abolish these components by means of drugs (24). We have seen also the first steps in the development of a mathematical model relating the probability of response to the intensity of the stimulus (32). At the auditory cortex, there have been attempts to present a mathematical description of repetitive responses (12) and to quantify the representations of the two ears (38).

RESULTS

The Round-Window Response to Single Clicks

In studying the click responses recorded with electrodes near the round window, we soon encounter some difficulties in the analysis of the electrical events. Ever since Davis and his group studied responses to clicks, investigators in this area have tried to differentiate between the so-called microphonic components and the action potentials or neurals. One of the first recognized advantages of the click stimulus was that under certain circumstances it gave rise to a response that could be analyzed into those two components. In Fig. 2 the electrical potentials recorded in response to weak clicks are preponderantly *neural* in origin. As the click gets more intense the neural potentials increase in size, but at intensities above −60 db another component, very different in character and behavior, makes its appearance. This is a so-called cochlear (or aural) microphonic; as the stimulus becomes more intense the microphonic increases rapidly in size and finally seems to swamp the neural components. Various methods have been used to detect the "neural signals" in the "microphonic noise."[4]

---

[4] Recently we have shown (35) that there exist electrode locations outside the bony shell of the cat's bulla at which we can record the neural response even to strong clicks without much interference from the microphonic. Others who are more concerned with microphonic potentials have taken a different view and have tried to get rid of the "neural distortion." However, this is not the place to review these methods or to discuss the existence of a possible third component—the so-called summating potential—recently described by Davis *et al.* (15).

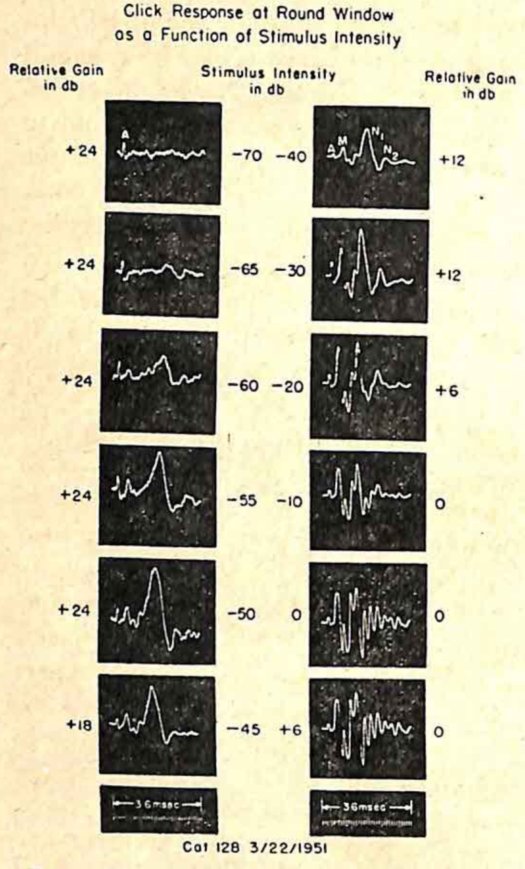


FIG. 2. The growth of the various components of the click response as seen by a wire electrode located at or near the round window.

For relatively weak clicks only the neural components (N_1 and N_2) can be detected in the background "noise." As the intensity of the click increases, microphonic activity becomes detectable ahead of the neural components: M is the earliest microphonic activity. A further increase in stimulus intensity brings out microphonic components that overlap in time with N_1 and N_2. For our strongest clicks (0 db and +6 db) distortion, shown by a flattening of the peak of M, is detectable. Note: *(a)* the latency of M is independent of stimulus intensity; *(b)* the latency of the peaks of the neurals decreases with intensity. (The artifact A indicates the time of delivery of the electrical pulse to the earphone.)

It is clearly beyond the scope of the present paper to discuss the origins and relations of the microphonic and neural potentials. The most commonly accepted hypothesis attributes the microphonic potentials to the organ of Corti (more specifically to the hair cells), whereas N_1 is supposed to be the summated response of the bipolar cells in the spiral ganglion and N_2 is imputed to neural structures in the cochlear nucleus. However, none of these statements has been established with sufficient certainty.

We will have to be satisfied with giving a rather gross description of how microphonics and neurals vary with stimulus intensity, that is, how the amplitudes and latencies behave as the intensity of our click is varied through a range of about 80 db.

If we refer to Fig. 2 we can identify a stimulus artifact (A), which occurs at the time at which the electrical pulse is delivered to the earphone. It takes the acoustic click about 0.3 msec. to travel from the earphone diaphragm to the cat's eardrum. Almost immediately thereafter, i.e., less than 0.1 msec. after the arrival of our acoustic click at the eardrum, an upward or downward deflection (M) is detectable for clicks that are 20 or more decibels above the animal's threshold.

The behavior of the microphonic. On most of our records M is a most prominent deflection in the complex of microphonic potentials. We have seen in Fig. 1 that the acoustic pulse lasts approximately 1 msec. If we make our click intense enough, however, we find that there are microphonic potentials present as late as 4 or 5 msec. after the delivery of the stimulus. At present we are inclined to attribute this prolonged microphonic activity to three factors: (*a*) the behavior of the cat's middle ear, which is vibrating at its characteristic frequency, (*b*) the fact that it takes a pressure wave a certain length of time to travel from the stapes to the helicotrema, and (*c*) the fact that, once a particular location of the cochlear partition has been displaced, it does not instantaneously return to its equilibrium position (4). For weaker clicks the late microphonic components are not in evidence since they are not sufficiently large in amplitude to stand out above the noise; it might, however, be possible to observe them even for weaker stimuli, were the electrode placed closer to the point of origin of these particular microphonics.

A few more considerations concerning M: The latency of M is independent of stimulus intensity. M reverses its polarity when a "rarefaction click" (sudden withdrawal of the earphone diaphragm as seen from the eardrum) is delivered instead of a "condensation click." There is evidence that this reversal in polarity of the microphonic potential holds, not only for M, but also for the later microphonic deflections. Such evidence can be obtained in a clear-cut manner only when we have either abolished the neurals or temporarily separated them from the microphonics (as in cooling, 29). As Fig. 2 shows, the amplitude of M grows with stimulus intensity up to a point where we encounter nonlinearity and overloading of either biomechanical or bio-electrical origin. The round-window response to the most intense clicks in Fig. 2 shows this phenomenon as a flattening or clipping of the peak of M.

Neural components. The neural components of the round-window response represent, in a sense, the way in which the click signal has been "coded" at the periphery of the auditory nervous system. We have already said that for weak clicks it is the neural components that are the most readily observed. As a matter of fact, when we talk about the "threshold" of an animal's response to clicks, we mean the stimulus intensity at which we can visually detect (on the oscilloscope face) neural responses on about half the trials. As can be seen from Fig. 2, neural components (N_1 and N_2 in their temporal order of appearance) stand out in the click response. Both these components are chiefly negative deflections, whether the stimulus is a "condensation click" or a "rarefaction click"; there are, however, changes in the wave form, the amplitude, and even the latency[5] of the neurals as we change the electrical polarity of our click.

As the intensity of the stimulus increases, the neurals grow and their latency decreases. A further increase in stimulus intensity leads to the previously mentioned contamination of the neurals by microphonics. Hence, it becomes difficult to give precise data on the behavior of the neurals for intense stimuli. There are statements in the literature to the effect that the neurals cease to grow in amplitude and stop decreasing in latency at about 30 db above threshold. We have, however, both direct and indirect evidence to the effect that the amplitude of the neurals keeps on growing and that the latency of both the onset and the peak of N_1 and N_2 decreases throughout the intensity range we have used.[6]

The Round-Window Response to the Second of a Pair of Clicks

Now let us see what happens when we

---

[5] We are dealing here with latency of a population of neural units. While most workers in physiology customarily define latency as the time interval between the stimulus and the first reliably observable deflection, we prefer for two reasons to refer to the latency of the peak of the neurals: (*a*) It is much easier to measure this latency since one is not as likely to be fooled by fluctuations in the base line. (*b*) The latency to the peak represents, more or less, a central tendency (mode) in the latency spectrum of the contributing neural units, and seems, therefore, the most representative measure.

[6] W. A. Rosenblith and M. R. Rosenzweig. Unpublished data.

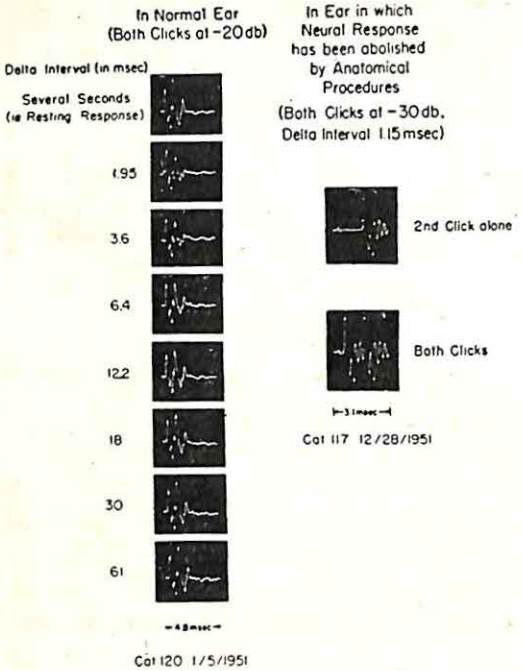


FIG. 3. Round-window responses to pairs of clicks.

In the left column are presented the responses to the second of a pair of clicks (the response to the first click is not shown; however, the tail end of the first response can be noted in the second trace from the top, where the delta interval is 1.95 msec.). As can be seen, M is not affected by the response to the preceding click whereas N$_1$ and N$_2$ are reduced in size. This reduction in size is the more noticeable, the shorter the delta interval. The effect is thus entirely confined to the neural components. This fact is emphasized by the traces in the right-hand column. These traces come from a cat in whose round-window response no neural components were detectable. Even for as short a delta interval as 1.15 msec. the responses to both members of a pair of clicks are identical. Presumably, these responses represent exclusively microphonic activity.

present two clicks, spaced apart by various time intervals, to the same ear of an anesthetized cat. (It will be convenient to refer to the first stimulus as S1 and the second stimulus as S2. The corresponding responses will be designated R1 and R2. The time interval between S1 and S2 will be called the delta interval.) The left column of Fig. 3 illustrates the phenomenon. The upper trace shows Ro, the resting or control response. It has a large microphonic deflection followed by two neural deflections. R2 for various delta intervals is shown in the traces below. We see clearly that some of the later deflections in R2 have been modified systematically; the shorter the delta interval, the less R2 resembles Ro. In an analogous experiment, the delta interval was held fixed and the intensity of S1 was altered (32). In that case, the more intense was S1, the less R2 resembled the control response. Analysis of the traces shows that the components affected by the preceding click are neural (N$_1$ and N$_2$); the microphonics remain unimpaired. Even the later microphonics are not affected, as the right-hand column of Fig. 3 demonstrates. (Cat 117 had been operated on by Dr. G. Rasmussen to eliminate the neural connections to the ear.) Here the response to a single click contained only microphonics. When two equally intense clicks were presented, R2 was indistinguishable from R1. All the evidence collected by us (and we have examined the round-window response to pairs of clicks in about 20 cats) points the same way. The modification in R2 amounts to a reduction in size of N$_1$ and N$_2$; M and the later microphonic components in R2 remain unaffected throughout the range of intensity we have used for S1 (about 80 db). We can, of course, easily imagine that a very intense S1 might produce enough damage in the inner ear to give rise to an effect on M in R2; however, we are not concerned here with such traumatic effects.

Let us now quantify further the effects of a preceding click upon the neural components in R2. We will measure the peak-to-peak amplitude of N$_1$, which we assume to be roughly proportional to the summated activity of a population of peripheral neural structures.[7] Figure 2 shows that the peak-to-peak amplitude of N_1 can be measured with better than 5 per cent accuracy for clicks that are about 20 to 30 db above the animal's "threshold." Though N_2 behaves qualitatively the same as N_1 (27), its measurement is not quite so easy at our round-window location.

Our recovery curves present the relative amplitude of N_1 or, in other words, the ratio of the peak-to-peak amplitude of N_1 in R2 divided by its amplitude in Ro. Figures 4, 5, and 10 show some of our data. We have chosen these particular animals to illustrate our findings because we have a sufficient number of observations on them under a variety of stimulus conditions, hence, reasonably small variability and rather clear-cut results. We have previously (39, Fig. 3) published recovery curves for equal intensities of S1 and S2. Such a choice of stimulus intensities narrows the range of observed phenomena unnecessarily and makes it much harder to get a picture of an over-all mechanism.

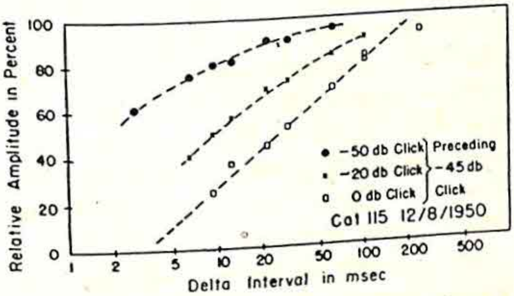


FIG. 4. Recovery of the amplitude of N_1 as a function of the delta interval (with the intensity of the first click as the parameter).

The relative amplitude of N_1 is defined as the ratio of the amplitude of N_1 in R2 divided by its resting amplitude (N_1 in Ro). The data presented here for a second click of constant intensity show that the amount of reduction in the relative amplitude depends upon both the intensity of the preceding click and the delta interval. Note that there is no supernormality in the recovery curve at the round window.

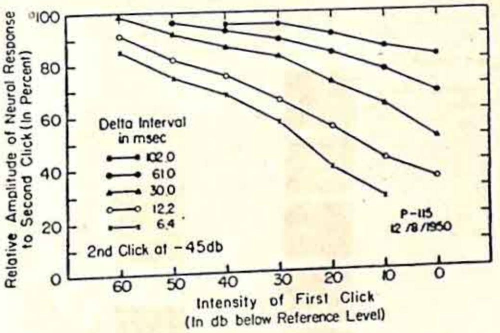


FIG. 5. Reduction of N_1 in R2 as a function of the intensity of S1 (with the delta interval as the parameter). For all delta intervals the relative amplitude of N_1 decreases with increasing intensity of the first click.

Figures 4 and 10 show recovery curves for two different animals. Both figures contain curves for which S1 was both less intense and more intense than S2. In the data for both these animals recovery is faster when S1 is the weaker of the two stimuli. Figure 4 shows that for a given intensity of S2 the recovery curves order themselves according to the intensity of S1. Figure 5 emphasizes this point more strongly by showing, for a given delta interval, the increasing effect upon N_1 (in R2) as the intensity of S2 is increased. The agreement in the general behavior of these recovery curves has already been stressed elsewhere (32). Suffice it to say that, if we make the necessary adjustments for different animals of the sensitivity constant (the statistical parameters representing the central tendency and the variance of the neural units), we can expect to obtain very close to the same recovery function for equivalent stimulus conditions. In other words, we may look forward toward a mathematical specifica-

---

[7] Experimentation dealing with the excitability cycle or with "refractoriness" has largely been concerned with the behavior of single neural elements. Let us stress here again that our gross electrodes recorded some sort of space and time averages of the action potentials of populations of neural elements.

tion of generalized recovery curves at the round window.

The delta interval covered by most of our recovery curves extends from about 6 msec. to about 200 msec. Data for shorter time intervals are hard to obtain because of the duration of R1: one gets temporal overlap between R1 and R2, a condition that makes it difficult, if not impossible, to measure the amplitude of N_1 with accuracy. We have, however, isolated observations under rather special stimulus conditions for delta intervals down to 1 msec. and even less. For all these time intervals there is still a depression in the amplitude of N_1 in R2. Also there is sometimes an increase in the latency of N_1 in R2. We are at present reserving these short delta intervals until we are ready to consider the latency problem; at that time we should also be prepared to consider whether the amplitude of the spikes contributed by the individual units has fully recovered. Note in this connection that the recovery curve on Fig. 4 for S1 = 0 db and S2 = −45 db (the lowest curve) intersects the time axis at a value appreciably greater than 1 msec., a figure generally given as the refractory period for the auditory nerve (14).

At the other end of our recovery curves the question of the length of the interval for complete recovery is not easily answered. If we plot the recovery curve on a linear time scale, we are effectively asking: when does the recovery curve reach the asymptote, RA = 100 per cent? If we use a device that might have pleased Verworn (43), namely, plotting recovery versus log *t*, we find what looks like a straight line when S1 is a strong stimulus and S2 a reasonably weak one. Under those circumstances S2 presumably samples from the same population that previously made up R1. Our confidence in both intercepts of the recovery curve with the 0 per cent and the 100 per cent line is still rather limited since (*a*) a small amount of variability in our experimental data may have rather important effects, (*b*) such physiological factors as temperature might be extremely critical at both ends of the curve, and (*c*) we are not necessarily dealing with a single straight line (i.e., a single recovery process), but possibly with two of slightly different slopes.

In summary, our recovery curves at the round window show several important features: (*a*) they are monotonic; (*b*) they never get to be supernormal, i.e., N_1 in R2 is never greater than N_1 of the resting response; (*c*) recovery is rapid for short time intervals (it may, however, take as long as 100 msec. and more to get back to 100 per cent relative amplitude); and (*d*) the particular shape of a recovery curve depends upon the intensities of S1 and S2 and the sensitivity constants of the animal. If S2 is a moderately weak stimulus and S1 is intense, the recovery curve will be very close to a straight line when plotted against log time.

The Cortical Response to Single Clicks

At the auditory area of the cortex we find responses considerably longer in duration and more varied in form than their progenitors at the round window. The cortical responses differ markedly in amplitude and form as we change the following factors: (*a*) the location of the electrode on the surface of the cortex, (*b*) the intensity of the stimulus, and (*c*) the physiological condition of the animal. Figure 6 shows some of the varieties of cortical responses. Each column of the figure presents responses obtained at a different location; at each of the three locations a series of eight stimulus intensities was employed. Each trace starts at the left with the delivery of the click. (In columns 1 and 3 a stimulus artifact was recorded at A.)

The earliest components of the responses are several wavelets; these small deflections appear at "A" in the top block of column 2. They reflect activity occurring at subcortical centers of the auditory system (23, 35) and are recorded at the cortex under specifiable conditions because of the electrical spread of the potentials. The subcortical responses are often very small when recorded at the cortex, and they do not appear clearly in columns 1 and 3. The traces of columns 2 and 3 show surface-negative deflections (upward in our convention) like the one at "B." These early negative deflections tend to be especially prominent when the active electrode is located near the suprasylvian sulcus, and the indifferent electrode is located at the mouth. As Fig. 6 illustrates, these negative deflections often become relatively more prominent as the stimulus intensity is reduced, apparently because the overriding positive deflection declines more rapidly with intensity. Figure 6, column 1, also illustrates the responses of a location at which no negative deflection was found at any stimulus intensity. The most characteristic feature of many cortical responses is a large surface-positive deflection: "C" is an example. The positive deflection may attain an amplitude of several hundred microvolts; the latency from stimulation to the peak of this deflection is usually 10 to 15 msec. Columns 2 and 3 show that with the weaker stimuli some locations show no positive deflection at all. Finally, the primary cortical response may conclude with another negative deflection like the one at "D." This deflection, when it occurs, often becomes detectable at about the same threshold as the positive deflection; in column 2 the positive and the later negative deflections both appear at about —50 db; in column 3, both appear at about —30 db.

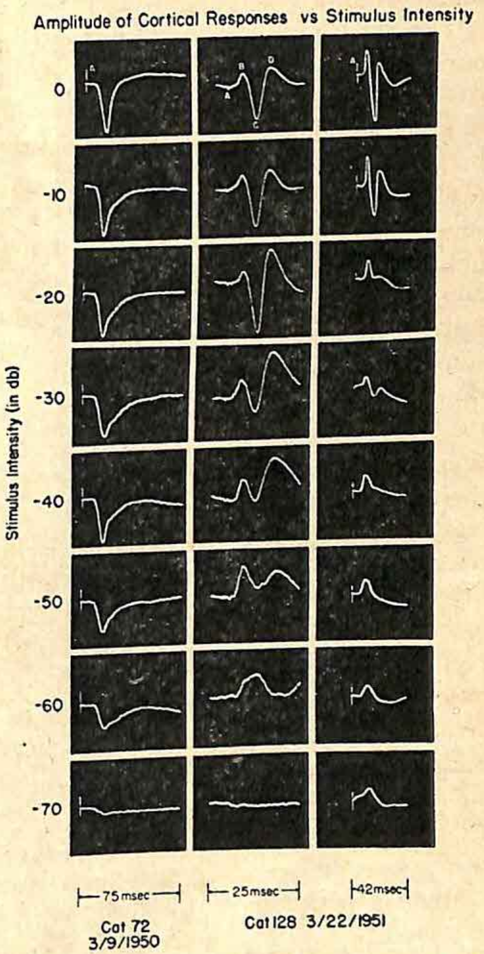


FIG. 6. Electrophysiological responses of the auditory cortex.

The three columns were recorded at three different electrode locations. The rows show responses to click stimuli of different intensities. The gain was constant in columns 1 and 3. In the middle column, the top two squares were recorded at the lowest gain; the next three squares had double this gain, and the bottom three squares had four times the gain of the top squares. Note the general increase in response amplitude with stimulus intensity. Differences among the wave forms of the responses are discussed in the text.

Not only is the primary response of

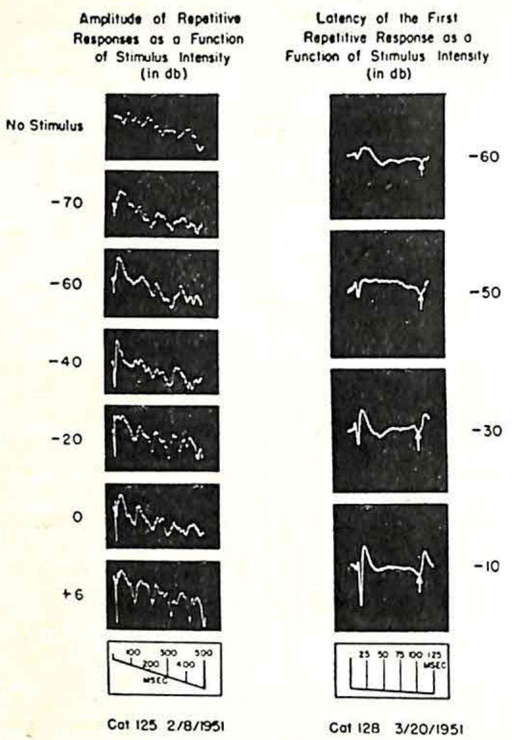


FIG. 7. Secondary responses of the auditory cortex.

The primary response at the extreme left of each trace is followed by the secondary responses. The left-hand column illustrates the increase in amplitude of the secondary responses with stimulus intensity; the right-hand column illustrates the decrease in latency of the secondary response with intensity.

the auditory cortex far longer in duration than the round-window response, but the stimulation may also evoke a train of secondary responses. Examples of secondary cortical responses are shown in Fig. 7. As compared to the primary responses the secondary responses are smaller in amplitude, longer in duration, and less regular in waveform. In some experiments the secondary responses resemble repetitions of the primary responses; in other experiments the secondary responses are relatively smooth sinusoidal undulations; in many cases there is little or no repetitive activity. The rate of repetition of the secondary responses is often similar to the rate of spontaneous activity observed in the same preparation. When we change the stimulus intensity the secondary responses exhibit alterations in amplitude and latency similar to those of the primary responses. Chang (11) has reported that the change in amplitude of the secondary responses is less prominent than the change in amplitude of the primary response; he has also reported that latency does not change over a wide range of intensities. We have found that the amplitude of the secondary responses often increases greatly as intensity is increased. Figure 7 illustrates this. The relative prominence of the change may depend upon the particular range of intensities studied. The amplitude of the secondary responses may continue to grow with increasing intensity even after the growth of the primary response has leveled off. We have also found that the latency of the secondary responses decreases as the stimulus intensity is increased. In the right-hand column of Fig. 7 the latency of the earliest secondary response decreases from 100 msec. at −60 db to 90 msec. at −10 db.

We have dealt with both the primary and secondary responses, but most of our measurements have been made on the surface-positive deflection of the primary response. It is important to note at the outset several obstacles to a rational quantification of this cortical response:

1. The observed primary response is the resultant of at least three potentials that overlap in time: an early surface-negative potential, a surface-positive potential, and a later surface-negative potential (these potentials are chiefly responsible, respectively, for deflections B, C, and D of Fig. 6). There is considerable evidence that the three potentials arise in different populations of cells.

Hawkins (23) reported that when he penetrated the cortex with an electrode the early negative potential remained unchanged but the later potentials reversed their signs. Bremer (8) concluded on the basis of a number of experiments that the surface-positive potential reflects the activity of cortical interneurons and that the later surface-negative potential reflects the activity of efferent pyramidal cells. Similarly, Bishop and Clare (5) have been able to assign the components of the response of the visual cortex to different cell populations.

The occurrence of these potentials overlapping in time means that the amplitude of the positive deflection does not reflect the positive potential alone. Our difficulties are illustrated in Fig. 8. Figure 8A is a highly schematic representation of three potentials; Fig. 8B shows the resultant deflection that is formed by algebraic addition of the three potentials. Exactly the same over-all deflection could have resulted from the addition of completely different sets of potentials. Figure 8C shows another set of potentials that adds up to the same resultant deflection. Furthermore, a change in the positive deflection may be caused chiefly by a change in a negative potential, as is illustrated in Figs. 8D and E. Here a decrease in the second negative potential, such as may occur under deep anesthesia, results in an increase in the positive deflection.

In quantifying the responses, the amplitude (peak voltage) has proved a more useful measure than the area (voltage times duration). The diagrams of Fig. 8 suggest why the amplitude of the positive deflection may represent the surface-positive potential more successfully than does the area. We have measured both amplitude and area under certain experimental conditions and have found the amplitude measurements to give more consistent results.

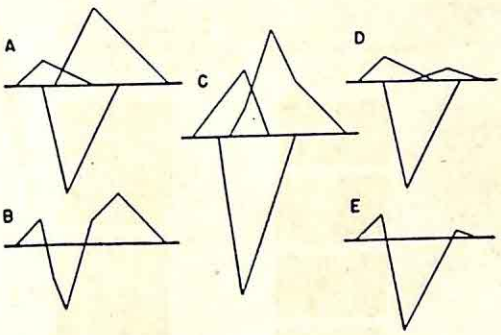


FIG. 8. Schematic representation of component potentials (A, C, D) and of composite cortical responses (B and E). Observed responses are the result of algebraic addition of several component potentials that overlap in time. The three potentials in A add up to give response B; the potentials in C also add up to response B; those in D, to response E. The waveforms of the component potentials cannot be inferred from the composite response for there is an infinite number of combinations of potentials that could result in the same composite response. These examples were chosen to illustrate the difficulties encountered when an analysis or breakdown of cortical responses into component potentials is attempted.

2. Under fixed experimental conditions successive responses may vary considerably. This variability is generally greatest when the level of anesthesia is light. To overcome some of the effects of this variability we have always recorded several successive responses for each experimental condition, and we will present the data in terms of the central tendency of the responses. In some cases, however, the distribution of response amplitudes is bimodal; hence it cannot be represented well by a central tendency.

3. Cortical responses vary in several respects as the recording electrode is shifted by as little as a millimeter across the surface of the auditory cortex. For example, the growth of response amplitude with stimulus intensity varies widely from location to location; a gen-

eral growth function cannot be given. Therefore, attempts to interrelate various functions (e.g., growth of R1 versus depression of R2, Fig. 13) can generally be done only with data obtained from the same cortical location.

The Cortical Response to the Second of a Pair of Clicks

In discussing the response of the auditory cortex to pairs of clicks we shall again designate the stimuli as S1 and S2, and the time interval between S1 and S2 as the delta interval; the primary cortical response to S1 will be called R1, and the primary response to S2, R2. Figure 9 shows responses to pairs of equally intense stimuli delivered to the same ear. The responses are abrupt downward (surface-positive) deflections; R1 occurs at the left of the trace, and R2 is displaced to the right in accordance with the delta interval. Successively larger delta intervals are shown in the successive rows.

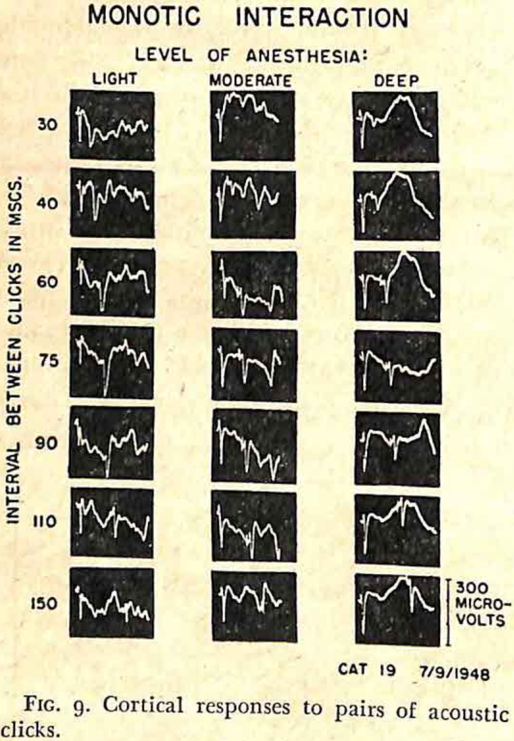


Fig. 9. Cortical responses to pairs of acoustic clicks.

The response to the first stimulus is the sharp downward deflection at the left of each trace. The response to the second stimulus is displaced to the right in accordance with the time interval between the two clicks. The amplitude of the first response (R1) remains relatively constant, but the amplitude of the second response (R2) varies systematically with the delta interval. Secondary responses appear in the middle column. (Reproduced from Fig. 14 of reference 37.)

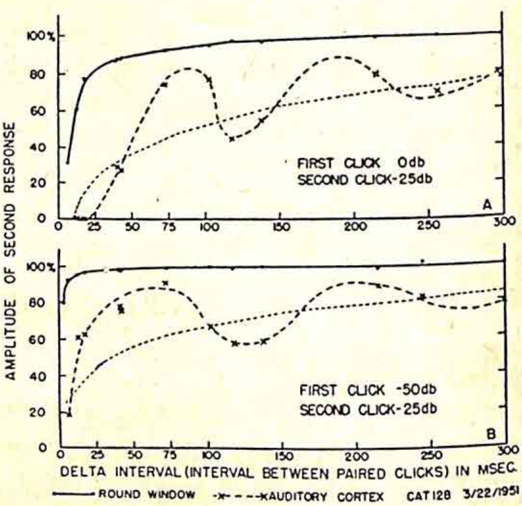


Fig. 10. Recovery curves at the round window and at the auditory cortex (data recorded simultaneously).

Note that recovery of the amplitude of N_1 is monotonic whereas the recovery curve for the cortical response seems to be made up of two components, one monotonic and the other cyclical. The monotonic component is indicated by the dotted line. The cyclical component can be obtained by subtracting the monotonic component from the cortical recovery function. In general, the recovery at the round window depends more strongly upon the intensity of the first click than does the recovery at the cortex.

Reading down any of the columns of Fig. 9, you see that the amplitude of R1 remains relatively constant, but the amplitude of R2 varies systematically as the delta interval increases. R2 grows over the first several intervals and then declines again. In column 1, R2 is depressed at an interval of 30 msec., enhanced at 60 and 75 msec., and depressed again at 110 and 150 msec. In the other columns the intervals at which depressed and

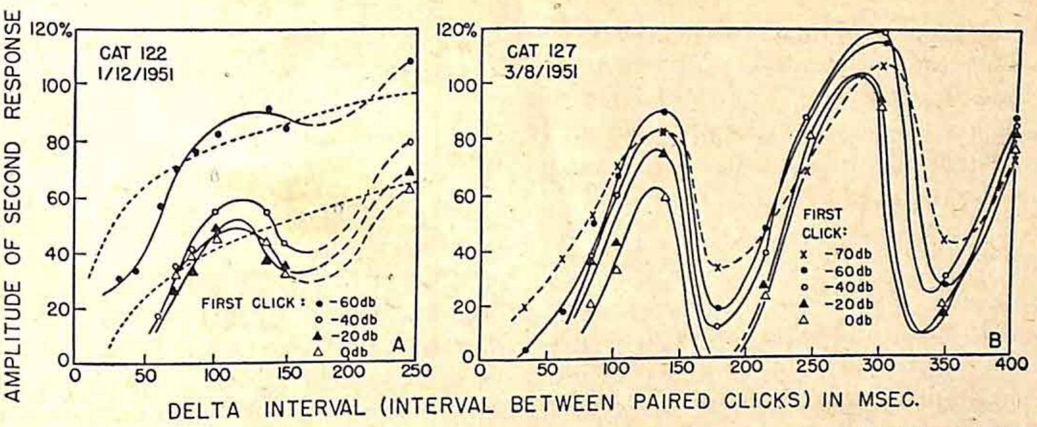


FIG. 11. Recovery of the relative amplitude of the second cortical response versus the delta interval (R2/Ro versus delta interval); the intensity of the first stimulus (S1) is the parameter. As S1 increases in intensity, the cyclical variations tend to become more pronounced and the recovery less pronounced at a given delta interval.

enhanced responses occur are different because of anesthetic effects, which will be discussed later. But every column demonstrates that the recovery of responsiveness at the auditory cortex is clearly not monotonic.

The course of recovery at the cortex is shown more clearly in the functions of Fig. 10. The cortical functions (dashed lines) show a marked alternation of peaks and troughs. This cyclical characteristic is also pronounced in Figs. 11 and 12. In an earlier study, Tunturi (42) presented a curve that suggested that recovery is monotonic in the geniculocortical radiations, but too few delta intervals were given to allow cyclical variations to appear. Jarcho (28), in experiments on recovery at somesthetic and auditory areas of the cortex, found "no curves suggesting the existence of supranormal or subnormal periods." With relatively light anesthesia his recovery functions showed "bizarre notching, and occasionally rather regular undulations." Chang (11) presented a recovery curve

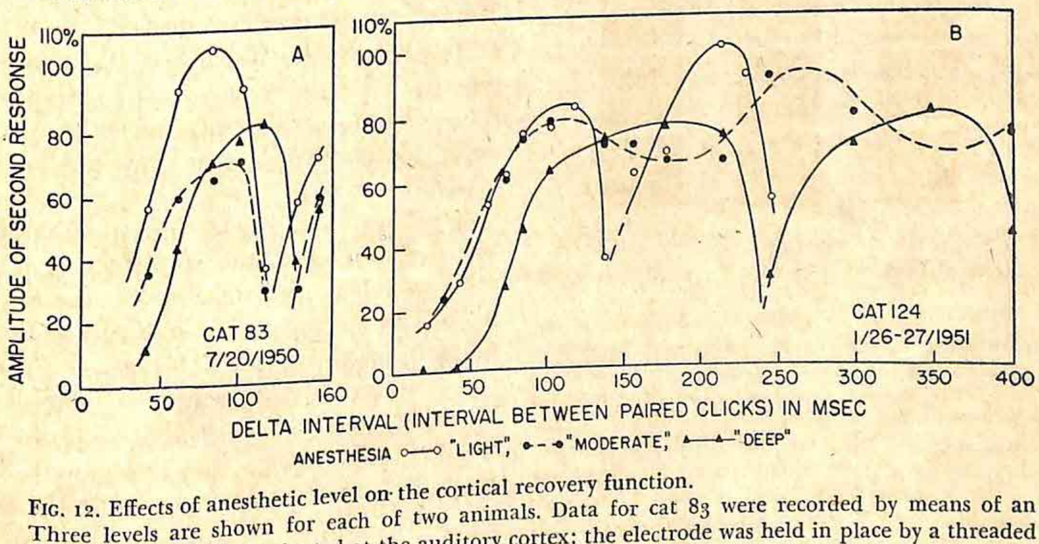


FIG. 12. Effects of anesthetic level on the cortical recovery function. Three levels are shown for each of two animals. Data for cat 83 were recorded by means of an electrode permanently implanted at the auditory cortex; the electrode was held in place by a threaded lucite button screwed into a tapped hole in the skull.

with small variations "suggestive of a periodic change in excitability of the cortex" following an auditory stimulus; when he replaced S1 by an electric shock to the cortex, the recovery curve showed large cyclical variations (12). Cyclical variations have also been reported in the recovery curves of the visual cortex (3, 6, 31) and of the somesthetic cortex (26). As we shall see, there are always cyclical alternations in the amplitude of R2 at the auditory cortex, but the prominence of the cycles depends upon the intensity of S1 and upon such physiological conditions as temperature and depth of anesthesia.

In the experiment of Fig. 10 simultaneous recordings were made at the round window and at the auditory cortex. The round-window functions (solid lines) show the monotonic and relatively rapid recovery that we have already described in an earlier section. It is apparent from Fig. 10 that the cortical recovery function differs strikingly from the round-window function. At the shorter delta intervals the cortical function may be influenced by a reduction of afferent input from the cochlea, but thereafter the cortical function reflects processes of only the higher centers of the auditory nervous system. The next sections of this report will be devoted to an analysis of the complex cortical recovery function. It will help to list first the main conclusions of this analysis.

1. The cortical recovery function can usefully be described as consisting of two components, one monotonic and the other cyclical. The monotonic component is indicated by the dotted logarithmic curves in Figs. 10 and 11. In each case, if this monotonic component is subtracted from the cortical recovery function, the cyclical component is obtained.

2. Although the two components are always present, they can in general be affected differentially by (*a*) varying the intensity of the first stimulus, (*b*) varying the level of anesthesia, and (*c*) varying the temperature of the animal.

3. The monotonic component of the recovery function is related to R1. This response represents the discharge of a population of cortical units; these units must recover their responsiveness before they can discharge again. Units other than those fired also become depressed in responsiveness following S1. The total initial depression of responsiveness is roughly proportional to the amplitude of R1. The recovery of responsiveness may follow the logarithm of the delta interval.

4. The cyclical component of the recovery function is an expression of the fluctuations in cortical responsiveness that follow R1. The secondary responses to S1 are also related to these fluctuations in responsiveness. The cyclical component of the recovery functions and the secondary responses exhibit behavior that is parallel in several respects.

The Monotonic and Cyclical Components of the Cortical Recovery Function

The simplest way to demonstrate the presence of two components in the recovery function is to vary the intensity of S1 systematically. The results of such experiments are shown in Fig. 11. In the experiment presented in the left-hand graph, S1 varied from 0 db to −60 db, while S2 was held at −20 db. When S1 was weak (60-db attenuation) the recovery function was nearly monotonic, though small cyclical variations could be seen. When S1 was made stronger, the cyclical component became more prominent. The amplitude of the cyclical component, measured in terms of complete recovery, was about 15 per cent when S1 was −60 db. It grew to about 30 per cent when S1 was increased to 0 db. The peak of the cyclical component came slightly earlier as the intensity of S1 was increased.

The general trend of the recovery function was also affected by the intensity of S1. The greater the intensity, the less recovery we find at each delta interval. The general trend of the recovery function can be approximated by a monotonic curve, such as the dotted curves in Fig. 10. Similar dotted curves have been plotted in the left-hand graph of Fig. 11 to show the general trend of recovery for S1 at −60 db and −20 db.

The results presented in the right-hand graph of Fig. 11 are similar to those in the left-hand graph, though in this case the cyclical component is so prominent that it may obscure other aspects of the recovery curves. We conclude that varying the intensity of S1 has differential effects on the cyclical and monotonic components of the cortical recovery function. As we increase the intensity of S1 we observe the following effects: the cyclical component stands out more clearly, and the monotonic recovery is slowed down; the period of the cyclical component shortens slightly, and the interval required for complete recovery of the monotonic component lengthens markedly.

It should be remarked that the considerable duration of the cortical recovery function makes it inadvisable to crowd successive trials together. We spaced trials at least one second apart, and, when the duration of the recovery function warranted it, we used intertrial intervals of two seconds or more.

The Level of Anesthesia and the Two Components

Varying the level of anesthesia also differentiates between the two components of the recovery function (see Fig. 9). One effect of this procedure is to change the amplitude of the cyclical component. The amplitude is usually greatest under conditions of moderately light surgical anesthesia. When anesthesia is very light there is a great deal of irregular spontaneous activity, and it is difficult to demonstrate a regular cyclical component. When the level of anesthesia is made very deep, the amplitude of the cyclical component is reduced from the maximum. We have never seen it abolished, even under very deep anesthesia.

The temporal characteristics of both components of the recovery function are also affected differentially by varying the depth of anesthesia. Data from two experiments are presented in Fig. 12. In each experiment we recorded responses from an animal at three levels of anesthesia. It is clear that the length of the period of the cyclical component became greater as the depth of anesthesia was increased. In the left-hand graph of Fig. 12 the peaks of the successive functions came at intervals of about 80, 100, and 120 msec. In the right-hand graph, the first peaks came at about 100, 130, and 180 msec. At the somesthetic cortex a similar dependence of the period of the cyclical variations upon the level of anesthesia has been reported (26).

The monotonic component of the recovery function—in so far as it can be estimated from these data—seems also to be affected by varying the level of anesthesia. As the depth of anesthesia is increased, the peaks of the recovery function generally fall lower as well as later. Estimates from several experiments indicate that by deepening the level of anesthesia we can lengthen the interval required for complete recovery from as little as 100 to 200 msec. to as much as several seconds. Varying the level of anesthesia seems to have more effect upon the interval required for complete recovery of the monotonic component than it has upon the period of the cyclical component. The data are not yet sufficient to establish the form of the relation between the temporal characteristics of the two components.

The effects of anesthesia are reversible. When the level was allowed to become lighter, the "time constants" characterizing the recovery function decreased. Evidently the recovery function is a fairly sensitive index of the level of anesthesia.

Gastaut *et al.* (20) have recently reported the effects of several anesthetics on the "cortical excitability cycle." The effects of the barbiturates were similar to those reported here. The effects of the other drugs can probably also be described in terms of our monotonic and cyclical components.

Temperature and the Two Components

A third procedure that affects the two components of the recovery function differentially is to vary the temperature of the experimental animal. Barbiturate anesthesia makes it difficult for the animal to maintain his temperature. Consequently the body temperature may be allowed to drop, or it may be kept at a particular level by application of heat. The effects of reducing the temperature are in general the same as those of deepening the anesthesia: the amplitude of the cyclical component may be affected, and the "time constants" of both components are lengthened. Obviously the temperature of the animal must be carefully controlled in all experiments on the recovery function of the cortex.

The Monotonic Component of the Cortical Recovery Function

Thus far we have shown that the cortical recovery function can be described in terms of a monotonic and a cyclical component, and we have mentioned three experimental procedures that have differential effects upon the two components. Now we may note some of the more important characteristics of each component. Let us start with the monotonic component.

The monotonic component of the recovery function is related directly to R1. This response—or, more accurately, the component cortical potential—represents the discharge of members of a population of cortical units. These units must recover their responsiveness before they can discharge again. To this extent, the cortical situation is analogous to the one we found at the round window, and our original hypotheses about the monotonic component were similar to the hypotheses that account for the round-window data. However, we have not been able to describe the cortical data completely by such a formulation since the cortical data have features not found at the round window. These features peculiar to the cortex are revealed best when the data are plotted to show the relation between R1 and the recovery of R2.

Figure 13 shows the relations between R1 and R2, found in two experiments. The figure presents the data of Fig. 11 replotted with each delta interval represented by a separate curve. Intervals through the first peak of the recovery function are shown in the left-hand graph of each pair, through the second peak in the right-hand graphs. In both experiments the curves start with a relatively straight descending portion, indicating a depression of responsiveness proportional to R1. The linear descent is then interrupted, probably owing to the cyclical component. The form of the curve after the early descent varies from

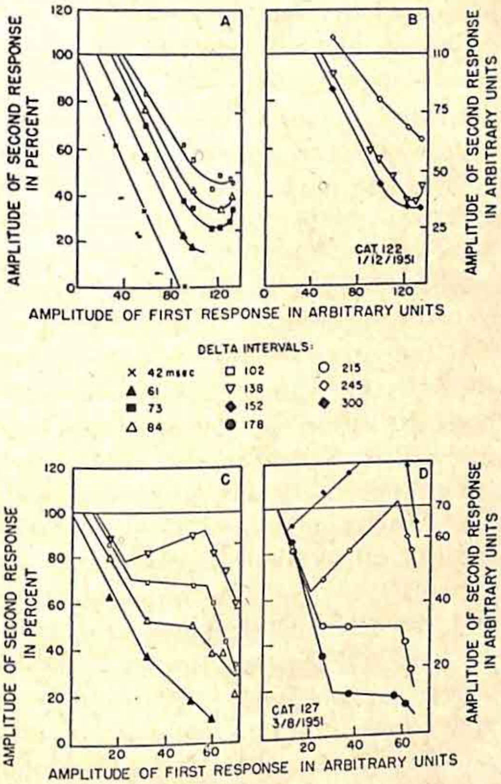


Fig. 13. Relative amplitude of the second cortical response versus amplitude of the first response (R2/Ro versus R1); the delta interval is the parameter. The same data are plotted on other coordinates in Fig. 11.

experiment to experiment. Two features of the early descent are of special significance and are characteristic of the cortical data; they mark important differences from the round-window data shown in Fig. 5:

1. The slope of the linear descent is slightly steeper than unity. That is to say, more cortical units are depressed in responsiveness at the time of R2 than were discharged to give R1.[8] This is an important difference from the round-window data, for there we used the depressed responsiveness to S2 as a measure of the excitation accomplished by S1; we assumed that the peripheral units became less responsive only as the result of having been discharged. At the cortex, however, we must conclude that some units became less responsive because of a stimulus without having contributed to the direct response to that stimulus. Amassian has reported such an effect at both the thalamus and the cortex (2). There are several ways in which this additional depression of responsiveness might come about: (*a*) The units might have discharged, but only tardily and asynchronously so that they did not enter into the R1 we recorded. (*b*) The units might have become less responsive because of changes of potential of neighboring units following the stimulus (9). (*c*) The units might have become less responsive when the discharge of other units decreased the facilitatory background activity (12). (*d*) The afferent tracts leading to the units might have become less responsive. Our data suggest that the first explanation cannot be the whole cause of the additional depression of responsiveness, but they do not allow us to differentiate among the other possibilities.

The steep slope is true for even the shortest delta intervals in the graphs, and data from other experiments demonstrate slopes steeper than unity when the delta interval is of the order of 1 msec. The processes underlying the depression of responsiveness must be rapid. They are more rapid than the "postexcitatory depression" which Chang (12) found in the motor cortex and also in the auditory cortex.

2. As the delta interval is increased, the slope of the lines decreases only slowly and the lines are displaced to the right. This is in marked contrast to the round-window data where the lines fan

---

[8] In block A of Fig. 13, the slope of R2 versus R1 is about — 1.2. When the amplitude of R1 is 0, the amplitude of R2 is 110; or R1 with an amplitude of only 90 can abolish R2 completely. In block C, the slope is about — 1.4.

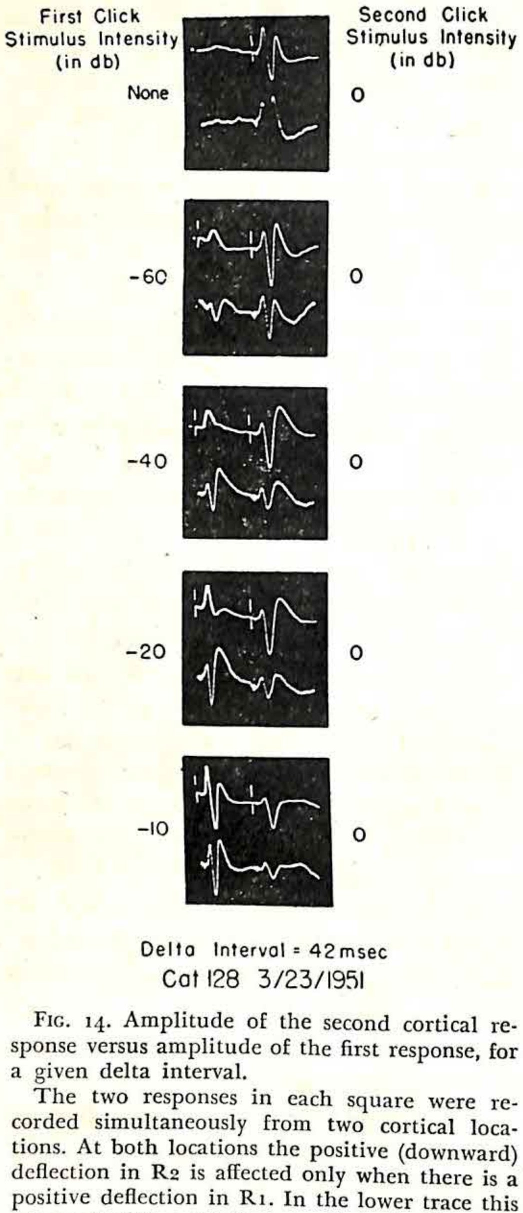


FIG. 14. Amplitude of the second cortical response versus amplitude of the first response, for a given delta interval.

The two responses in each square were recorded simultaneously from two cortical locations. At both locations the positive (downward) deflection in R2 is affected only when there is a positive deflection in R1. In the lower trace this occurred with a relatively weak (—60 db) click; in the upper trace it required a relatively strong (—20 db) click.

out widely from a more or less common origin as shown in Fig. 5. The displacement to the right in the cortical data means that at the given delta interval there may be complete recovery following a small R1, but incomplete recovery following a larger R1. In other words, recovery may proceed more rapidly following a small R1 than a large R1.

It may be asked why we have plotted the recovery function in the form R2 versus R1 at the cortex (Fig. 13), when we plotted it as R2 versus S1 at the round window (Fig. 5). At the cortex R2 versus S1 does not give consistent results, probably because the function relating amplitude of response to stimulus intensity varies so widely from location to location. Figure 14 demonstrates how awkward it would be to plot R2 versus S1 at the cortex. The two traces in each block were recorded simultaneously from two cortical locations in the same preparation. We varied the intensity of S1 systematically while holding S2 and the delta interval constant. In the case of the upper trace, no sizable positive deflection occurs in R1 until —10 db, and the amplitude of R2 is not importantly affected before this point. In the lower trace, on the contrary, a positive deflection is already present in R1 at —60 db, and the R2 for this location is already affected at —60 db. Plotting R2/Ro versus S1 for different cortical locations would obviously give widely discordant functions; plotting R2/Ro versus R1 in a number of experiments has given relatively linear functions (over the first part of the plot), though the slope varies somewhat from location to location.

At the round window the function relating the response amplitude of N_1 to stimulus intensity is monotonic and may be described by the normal probability integral (32); thus the recovery function can be plotted as R2/Ro versus S1 with consistent results. On the other hand, it is difficult to plot R2 versus R1 with round-window data because the range of stimulus intensity over which R1 can be measured without microphonic interference is relatively small. Moreover, when S1 is weak and S2 is of medium intensity, R2 is appreciably affected only at the shortest delta intervals. In spite of these difficulties we have attempted to plot N_1 of R2 versus N_1 of R1. The resulting functions, in so far as we can determine them, seem to be consistent with Fig. 5 and with the mathematical formulation of round-window data to which we referred earlier.

Let us note briefly that the recovery functions need not be the same for all deflections of the cortical response. In the upper traces of Fig. 14 it is apparent that the initial negative deflection of R2 declines steadily as the initial negative deflection of R1 grows; the later negative deflection (D) of R2 meanwhile declines slightly and changes its form; the positive deflection of R2 does not decline until the positive deflection of R1 grows. Here as in other experimental situations the initial negative deflection seems to behave rather independently of the later deflections.

The Cyclical Component of the Cortical Recovery Function

In this section we will complete our description of the cyclical component, and we will mention some ways in which the secondary cortical responses may exthe secondary cortical responses may exhibit parallel characteristics. We have already shown three characteristics of the cyclical component: (*a*) Its amplitude increases as the intensity of S1 is increased. (*b*) The length of its period decreases slightly as the intensity of S1 is increased. (*c*) The length of its period increases as the anesthesia is deepened or as the body temperature is lowered.

To this list we can add the three following characteristics: (*d*) The cyclical variation swings to both sides of the monotonic component. That is to say, the cyclical component is an alternation between enhanced and depressed responsiveness. The cortical recovery function may even become supernormal, as in Fig. 11; such supernormality is never seen in the responses of the round window. Previous reports have compared the cyclical alternations with the level of 100 per cent recovery rather than with the monotonic component, and this has led to apparently important discrepancies between results. Jarcho (28) found that "all deviations of the curve from the theoretical horizontal were downward, i.e., represented excitability depressed from the level predicted." On the other hand Chang (12) pictured the alternations as centering around the 100 per cent level. Actually more of his measurements were above 100 per cent than below it; presumably this occurred because his curve resembled the half-wave rectification of a sine wave, though Chang represented it schematically as a sine wave. Gastaut *et al.* (20) approached the concept of the monotonic component when they described their recovery functions in terms of the "average value" after the initial ascent. They found that although a few curves oscillated around the 100 per cent level after the initial ascent, the average value of many curves lay below the 100 per cent level, and the average value of a few curves lay above it. To account for these differences in their average values, they hypothesized long-lasting postexcitatory states of subnormality or supernormality. We can describe the results of Jarcho, Chang, and Gastaut *et al.*, except for the few cases with supernormal base lines, in terms of cyclical variation around a monotonic component. It is possible that the long-lasting supernormality hypothesized by Gastaut *et al.* may reflect only the nonsinusoidal shape of some curves in which more measurements lie above the axis than below it. (*e*) The form of the cyclical component varies widely. It may be smooth and almost sinusoidal, or it may be sharp and almost saw-toothed. Generally, the enhanced phase is relatively smooth, and it usually lasts more than half the cycle. The cyclical component may look less like a sine wave than like the full-wave rectification of a sine wave (see Fig. 12). (*f*) The cyclical component behaves roughly like a negative sine wave that starts ($t = 0$) at a delta interval of 0. The cyclical component starts with the development of the initial negative trough (depressed responsiveness). The first peak (enhanced responsiveness) usually occurs after about three-quarters of a period. The length of a period has, of course, to be estimated by measuring from one peak to the next or from one trough to the next. When the cyclical component is nearly saw-toothed in form, each peak is followed very quickly by a trough. In this case the first peak of the recovery function may occur almost a complete period after R1.

The cortical after-activity may show characteristics similar to those of the cyclical component. We wish to emphasize, however, that the cyclical component may be well defined even at a location, and under experimental conditions where there is little or no regular after-activity. The recovery function is a more direct and more stable measure of cortical responsiveness than is the after-activity. The after-activity (including the repetitive responses when they are present) parallels the cyclical component in the following respects: (*a*) It exhibits characteristics *a*, *b*, and *c* listed above for the cyclical component (see Fig. 7). (*b*) The developing surface-positive phases of the secondary responses coincide with the peaks of the cyclical component; the declining phases (return to the base line) coincide with the troughs of the cyclical component (12). (*c*) The wave forms of the after-activity and of the cyclical component tend to show some similarities. When the after-activity consists of rather smooth waves, the cyclical component tends to be rather sinusoidal. When there are large repetitive responses, the cyclical component tends to be rather saw-toothed.

Responses to Pairs of Dichotic Stimuli

The cortical phenomena we have described can also be demonstrated when S1 is delivered to one ear and S2 to the other (dichotic stimulation). The prominent binaural interaction observed at the cortex is illustrated in Fig. 15. At the round window, in marked contrast, we have not been able to find convincing evidence of binaural interaction, though there are sometimes misleading indications of interaction arising from leakage of acoustic stimuli or from spread of electrical potentials (35).

The relative amplitudes of the two cortical responses vary according to

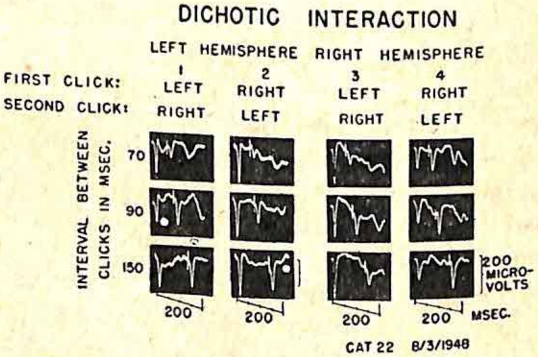


FIG. 15. Cortical responses to pairs of dichotic stimuli.

Even when the first and second stimuli are delivered to different ears, the second response (R2) is affected by the first response (R1). The order in which the ears were stimulated is shown by the sequences left-right or right-left above the columns. (From Fig. 10 of reference 37.)

which ear receives the first stimulus and which cerebral hemisphere the electrode is on. A stimulus to one ear alone tends to produce a larger response at the contralateral hemisphere (38). When the electrode is on the left hemisphere, stimulating the ears in the order left-right (column 1 of the figure) produces a smaller R1 and a larger R2 than stimulating them in the order right-left (column 2). At the right hemisphere the opposite effect is observed: here stimulation in the order left-right (column 3) produces a larger R1 and a smaller R2 than in the order right-left (column 4).

When the delta interval for dichotic stimuli is sufficiently brief, a single response occurs, and this response is often larger than the response to either one of the stimuli. The response tends to be larger at the hemisphere contralateral to the ear that receives the first stimulus. Thus, if we stimulate the ears in the order left-right, the response at the right hemisphere tends to be larger than the response at the left hemisphere. Significant differences can be observed even when the delta interval is well under 1 msec. (37). We have shown elsewhere (38) that the data obtained with dichotic stimulation can be interpreted in the following way: The two ears are represented at each cerebral hemisphere by two partially overlapping populations of cortical units; at each hemisphere the population that represents the contralateral ear is larger than the population that represents the ipsilateral ear.

DISCUSSION

One of the major difficulties in interpreting the responses of a population of neural units stems from the fact that the population may not be homogeneous. In the recovery functions of the round window, for example, there is evidence that the rates of recovery may not be homogeneous; there may be more than one process or population involved. At the auditory cortex it is clear that the rates of recovery are not homogeneous; at a delta interval at which recovery is already complete after a weak S1, there may still be depression after a strong S1. We may speculate that there are subpopulations such that those with higher thresholds have somewhat slower rates of recovery, or that larger populations as such have slower rates of recovery. We can suggest two procedures that may be useful in providing more homogeneous recovery curves and in determining processes of the auditory nervous system in greater detail.

1. The use of tone pips (15) or filtered clicks may permit the separate activation of subpopulations of units that have limited ranges of frequency sensitivity, and that may also be more homogeneous in other respects. The click stimulus is equivalent to a rather wide-band acoustic disturbance, and it tends to activate units sensitive to different frequencies. A tone pip or filtered click that excites a restricted subpopulation may be followed by a more homogeneous recovery curve.
2. Simultaneous recording with microelectrodes and gross electrodes may shed light on the relations between the activity of single units and the over-all response. This could help with the problem of inferring component cortical potentials from the composite response as well as with the inhomogeneity of functions of the over-all response. It should be pointed out that the study of the behavior of single units alone will probably not solve our problems concerning the over-all responses of the cortex. For one thing, the single units provide only a sample of the total population: exacting sampling techniques must be used if the recordings are to provide an adequate and representative sample of the total population. Second, the recorded over-all response is not the simple algebraic sum of the independent responses of single units. To obtain the over-all response, interaction of the units must be taken into account, and the unit responses must be weighted according to their distances from the gross electrode and according to other factors that determine the relative contribution of their potentials to the over-all response.

Terminology

Writing about the recovery functions at the round window and auditory cortex has posed problems of terminology. What is needed is a set of essentially descriptive terms that meet the following requirements: (*a*) the terms should be applicable primarily to populations of units since the recorded responses come from populations, but they should not be inapplicable to single units; (*b*) they should be capable of describing events occurring at any level of the nervous systems from peripheral to central locations; and (*c*) they should not specify or even imply processes underlying the observed events. We decided upon the terms "responsiveness," "depressed responsiveness," and "enhanced responsiveness" as being most nearly neutral and operational at the same time.

Many terms in current use cannot be employed to describe observations without suggesting that certain levels of the nervous system or certain physiological processes are involved. The following list gives some of these terms and the restricted connotations that are frequently attached to them: (*a*) *Refractory period*—a period during which the threshold of a neuron is raised following the conduction of a nerve impulse. The use of the term implies that the processes involved all occur within the individual neuron and that no interaction between neurons is involved. Refractory period has not often been used in connection with higher centers of the nervous system; it has been used (as by Elsberg and Spotnitz, 16) as a concept derived from the behavior of the whole organism. (*b*) *Subnormality*—raised threshold of a neuron referable to the existence of a positive after-potential following the conduction of a nerve impulse (19). (*c*) *Indirect inhibition*—failure of response to afferent impulses because of subnormality (17). (*d*) *Direct inhibition*—failure of response to afferent impulses because of the effects of simultaneously converging impulses (17). (*e*) *Excitability* and (*f*) *irritability* suggest threshold phenomena in individual neurons.

In treating cortical data it is obviously necessary to discriminate between descriptive and explanatory terms. We can describe much more than we can explain. For example, we observed an "additional depression of responsiveness" at the cortex; we noted four possible explanations for this phenomenon, but the present data do not permit us to commit ourselves to any one of the four. At the periphery it is equally important to discriminate between description and explanation.

When a general descriptive term such as "modified responsiveness" is used, care must be taken to specify whether it refers to a single unit or to a population, and what change in the response is meant. In the nerve, for example, the response may change in threshold, latency, amplitude, and conduction velocity, and each change may have a different course from the others (21, 36). At the round window and at the cortex we have quantified the changes that occurred in the amplitude of the responses. We could have quantified other aspects of the response, but the present results should not be applied a priori to any other aspects. We may note at the same time that our cortical observations were restricted to auditory areas I and II of the cortex. While some of our conclusions may be applicable to other sensory areas of the cortex, we have no evidence at present that they are.

Absolutely Unresponsive Period

A descriptive term that we have avoided is "absolutely unresponsive period." *Absolute* unresponsiveness can be defined only with extremely intense stimuli, which we did not employ.[9] Moreover, Rosenblueth *et al.* (36) question whether there is any evidence for the existence of an absolutely refractory period in nerves; they find only a "functionally refractory period" during which a nerve can respond but the response is not propagated. It has been reported that the auditory cortex of anesthetized animals shows an absolutely unresponsive period that may last for 8 msec. (11) or for as long as 80 msec. (28) or 100 msec. (42). However, unless the experiments cited employed extremely intense stimuli, the reports of *absolute* unresponsiveness should be questioned. In similar experiments we found that the cortex often responded to more intense stimuli during at least part of this so-called absolutely unresponsive period.

---

[9] The use of extremely intense stimuli in the intact animal would pose the problem whether such stimuli might not tend to produce irreversible changes in the peripheral part of the auditory system.

It is often asked how far apart in time two clicks must be spaced in order to be perceived as separate. An answer to this question depends on such factors as the intensities of the two clicks and the particular perceptual criteria used. The round-window recovery functions suggest a lower limit to the delta interval at which the second click can evoke a sizable response. The cortical recovery functions suggest that relatively long intervals are necessary for the second response to appear—longer in fact than perceptual experiments require (44)—but it must be remembered that the depression at the cortex is considerably prolonged by the use of anesthesia. The long duration of these cortical recovery functions also indicates why frequency-true responses to pure-tone stimuli cannot be expected at the auditory cortex of an anesthetized animal.

Temporal Characteristics

When the observed recovery function can best be described as the resultant of two or more components, it may be more convenient as well as more correct to assign temporal characteristics to the components instead of to the composite function. For example, it is difficult to describe the temporal characteristics of the recovery function observed at the auditory cortex, but the course of the function can be described as the resultant of the rate of recovery of the monotonic component (roughly proportional to the logarithm of the delta interval) and the period of the cyclical component (which follows roughly a negative sine function of the delta interval). The attempt in two recent papers to assign temporal characteristics to the composite cortical recovery function led to this rather arbitrary procedure: The end of the cortical "relatively refractory period" was taken as that point at which the recovery function crossed the 100 per cent level on its way to a peak of enhanced responsiveness. Besides helping to preclude such procedures, the attempt to describe accurately the temporal characteristics of the components of the recovery curve may help us to define more clearly the processes underlying the components.

SUMMARY

Electrical responses to a single acoustic click and pairs of clicks were recorded at the round window of the cochlea and at areas I and II of the auditory cortex in anesthetized cats. The behavior of these responses is described as a function of the intensities of the stimuli, the time interval between paired stimuli (delta interval), and the physiological state of the animal. At both locations the response to a second of a pair of clicks differs, for considerable delta intervals, from the resting response.

Round Window

Recovery curves for the so-called first neural component [$R2/Ro = f(t)$] show that (*a*) recovery is monotonic and there is no evidence for supernormality; (*b*) the recovery curves are negatively accelerated with the shape of a particular recovery cúrve depending upon the intensities of both clicks (it is, however, possible to predict the relative position of a family of parametric recovery curves); and (*c*) reference is made to a mathematical formulation, given elsewhere, that tries to account for the observed data on the basis of a probability model.

Auditory Cortex

The recovery functions of the auditory cortex all show a general monotonic rising trend around which cyclical variations occur. The monotonic and cyclical components of the recovery function can be affected differentially by varying: (*a*) the intensity of S1, the conditioning stimulus, (*b*) the level of anesthesia, and (*c*) the body temperature of the experimental animal.

The shape of the monotonic component is difficult to determine with precision in the presence of the cyclical component; it may follow the logarithm of the delta interval. The monotonic component reflects the recovery of responsiveness of the cortical units that were rendered unresponsive following S1, including units that became less responsive even though they did not contribute directly to R1. The depression of responsiveness following S1 is roughly proportional to the amplitude of R1, at least for the smaller values of R1. The maximum depression of responsiveness develops within a few milliseconds after R1; the rate of recovery varies with the intensity of S1 and with the physiological state of the preparation.

The cyclical component behaves roughly like a negative sine function that starts at a delta interval of 0. The cortical after-activity exhibits behavior parallel to that of the cyclical component in several respects, but the cyclical component may be prominent in a case where little or no after-activity can be seen.

The recovery function of the auditory cortex differs importantly from that of the round window in several respects. The cortical function, but not the round-window function, shows (*a*) the presence of a cyclical component; (*b*) the presence, in many cases, of one or even several supernormal periods; (*c*) longer duration of the depression; (*d*) greater sensitivity to changes of anesthetic level and temperature; (*e*) spread of depression to units that did not contribute, at least directly, to R1; (*f*) marked inhomogeneity of rates of recovery depending upon the intensity of S1; and (*g*) depression following dichotic as well as monotic stimulation.

Difficulties of interpreting and analyzing the composite responses recorded with gross electrodes were noted in regard to the overlapping component cortical potentials and to the inhomogeneity of recovery functions.

REFERENCES

1. Ades, H. W., & Brookhart, J. M. The central auditory pathway. *J. Neurophysiol.*, 1950, 13, 189-206.
2. Amassian, V. E. Sensory interaction. *Fed. Proc.*, 1951, 10, 6.
3. Bartley, S. H. *Vision, a study of its basis.* New York: Van Nostrand, 1941.
4. Békésy, G. v. Über die Resonanzkurve und die Abklingzeit der verschiedenen Stellen der Schneckentrennwand. *Akust. Z.*, 1943, 8, 66-76.
5. Bishop, G. H., & Clare, M. Radiation path from geniculate to optic cortex in cat. *J. Neurophysiol.*, 1951, 14, 497-505.
6. Bishop, G. H., & O'Leary, J. S. Electrical activity in the lateral geniculate of cats, following optic nerve stimuli. *J. Neurophysiol.*, 1940, 3, 308-322.
7. Bremer, F. Études oscillographiques des résponses sensorielles de l'aire acoustique corticale chez le chat. *Arch. Int. Physiol.*, 1943, 53, 53-103.
8. Bremer, F. Les réactions auditives de l'écorce. *L'année psychol.*, 1951, 50, 115-128.
9. Brooks, C. M., & Eccles, J. C. An electrical hypothesis of central inhibition. *Nature, London*, 1946, 159, 760-771.
10. Buytendijk, F. J. J., & Meesters, A. Duration and course of the auditory sensation. *Commentationes Pontif. Acad. Sci.*, 1942, 6, 557-576.
11. Chang, H.-T. The repetitive discharge of the corticothalamic reverberating circuit. *J. Neurophysiol.*, 1950, 13, 235-258.
12. Chang, H.-T. Changes in excitability of cerebral cortex following single electric shock applied to cortical surface. *J. Neurophysiol.*, 1951,14, 95-112.
13. Davis, H. The electrical phenomena of the cochlea and the auditory nerve. *J. acoust. soc. Amer.*, 1935, 6, 205-215.
14. Davis, H. Psychophysiology of hearing and deafness. In S. S. Stevens (Ed.), *Handbook of experimental psychology*. New York: Wiley, 1951. Pp. 1116-1142.
15. Davis, H., Fernandez C., & McAuliffe, D. R. The excitatory process in the cochlea. *Proc. nat. Acad. Sci.*, 1950, 36, 580-587.
16. Elsberg, C. A., & Spotnitz, H. The relative refractory period of olfaction and vision. *Bull. Neurol. Inst., N. Y.*, 1938, 7, 78-94.
17. Fulton, J. F. *Physiology of the nervous system.* New York: Oxford Univer. Press, 1943.
18. Galambos, R., Rose, J. E., Bromiley, R. B., & Hughes, J. R. Microelectrode studies on medial geniculate body of cat. II. Response to clicks. *J. Neurophysiol.*, 1952, 15, 359-380.
19. Gasser, H. S. Conduction in nerves in relation to fiber types. *Res. Publ. Ass. nerv. ment. Dis.*, 1935, 15, 35-59.
20. Gastaut, H., Gastaut, Y., Roger, A., Corriol, J., & Naquet, R. Étude électrographique du cycle d'excitabilité cortical. *EEG clin. Neurophysiol.*, 1951, 3, 401-428.
21. Graham, H. T., & Lorente de No, R. Recovery of blood-perfused mammalian nerves. *Amer. J. Physiol.*, 1938, 123, 326-340.
22. Gross, N. B., & Thurlow, W. R. Microelectrode studies of neural auditory activity of cat. II. Medial geniculate body. *J. Neurophysiol*, 1951, 14, 409-422.
23. Hawkins, J. E., Jr. Electrophysiology of the auditory cortex. Unpublished doctoral dissertation, Harvard Univer., 1941.
24. Hawkins, J. E. Cochlear signs of streptomycin intoxication. *J. Pharmacol exp. Therapeutics*, 1950, 100, 38-44.
25. Hawkins, J. E., Jr., & Kniazuk, M. The recovery of auditory nerve action potentials after masking. *Science*, 1950, 111, 567-568.
26. Heinbecker, P., & Bartley, S. H. Action of ether and nembutal on the nervous system. *J. Neurophysiol.*, 1940, 3, 219-236.
27. Heise, G. A., & Rosenblith, W. A. Electrical responses to acoustic stimuli recorded at the round window of the pigeon. *J. comp. physiol. Psychol.*, 1952, 45, 401-412.
28. Jarcho, L. W. Excitability of cortical afferent systems during barbiturate anesthesia. *J. Neurophysiol.*, 1949, 12, 447-457.
29. Kahana, L., Rosenblith, W. A., & Galambos, R. Effect of temperature change on the round window response of the hamster. *Amer. J. Physiol.*, 1950, 163, 213-223.
30. Kemp, E. H., Coppée, G. E., & Robinson, E. H. Electric responses of the brain stem to unilateral auditory stimulation. *Amer. J. Physiol.*, 1937, 120, 307-315.
31. Marshall, W. H. Excitability cycle and interaction in geniculate-striate system of cat. *J. Neurophysiol.*, 1949, 12, 277-288.
32. McGill, W. J., & Rosenblith, W. A. Electrical responses to two clicks: a simple statistical interpretation. *Bull math. Biophysics*, 1951, 13, 69-77.
33. Rawdon-Smith, A. F., & Hawkins, J. E., Jr. The electrical activity of a denervated ear. *Proc. roy. Soc. Med.*, 1939, 32, 496-511.
34. Rosenblith, W. A. Auditory masking and fatigue. *J. acoust. Soc. Amer.*, 1950, 22, 792-800.
35. Rosenblith, W. A., & Rosenzweig, M. R. Electrical responses to acoustic clicks: influence of electrode location in cats. *J. acoust. Soc. Amer.*, 1951, 23, 583-588.

36. ROSENBLUETH, A., ALANIS, J., & MANDOWKI, J. The functional refractory period of axons. *J. cell. comp. Physiol.*, 1949, **33**, 405-439.

37. ROSENZWEIG, M. R. Electrophysiological correlates of auditory localization. Unpublished doctoral dissertation, Harvard Univer., 1949.

38. ROSENZWEIG, M. R. Representations of the two ears at the auditory cortex. *Amer. J. Physiol.*, 1951, **167**, 147-158.

39. ROSENZWEIG, M. R., & ROSENBLITH, W. A. Some electrophysiological correlates of the perception of successive clicks. *J. acoust. Soc. Amer.*, 1950, **22**, 878-880.

40. STEVENS, S. S., & DAVIS, H. *Hearing, its psychology and physiology*. New York: Wiley, 1938.

41. THURLOW, W. R., GROSS, N. B., KEMP, E. H., & LOWY, K. Microelectrode studies of neural auditory activity of cat. I. Inferior colliculus. *J. Neurophysiol.*, 1951, **14**, 289-304.

42. TUNTURI, A. R. A study on the pathway from the medial geniculate body to the acoustic cortex in the dog. *Amer J. Physiol.*, 1946 **147**, 311-319.

42. VERWORN, M. *Irritability*. New Haven: Yale Univer. Press, 1913.

44. WALLACH, H., NEWMAN, E. B., & ROSENZWEIG, M. R. The precedence effect in sound localization. *Amer. J. Psychol.*, 1949, **62**, 315-336.

(Accepted for publication May 28, 1953)

Vol. 67, No. 14 Whole No. 364, 1953

Psychological Monographs: General and Applied

Job Satisfaction as Related to Need Satisfaction in Work

Robert H. Schaffer[1]
Douglas Williams Associates, New York

I. INTRODUCTION

THE PRESENT study is an attempt to develop a theoretical conceptualization of job satisfaction which will have functional utility.

Work as a psychological function has been the concern of investigators in many fields for the past thirty or more years. The early endeavors by psychologists as well as by researchers of other disciplines were chiefly directed at determining relationships between various psychophysical and physical characteristics and productivity. Light, noise, and fatigue are typical of the variables selected for investigation. It was more or less natural that the early research work in this field should take this course since much of the impetus was at the time provided by the "scientific management" movement identified with men like Taylor and Gilbreth. However, the insights provided by the Hawthorne experiments (11), which began in 1926, stimulated a shift of emphasis from the physical to the emotional aspects of worker behavior.

Beginning in the 'thirties a large number of social scientists set their sights at understanding the human factors operating in the work environment. Many of them were specifically interested in the factors related to worker adjustment and satisfaction. That psychologists and workers in allied fields have devoted considerable energy to attacks on this problem is attested to by the numbers and variety of research efforts periodically reported by Hoppock (4).

In reviewing the work in this area one has the feeling that the progress made has not been commensurate with the great effort expended. One of the reasons for this is thought to be the specific and empirical nature of the studies conducted. Very few workers have shown a great interest in the theoretical aspects of the problem; most have been content to investigate specific variables without considering how these variables fit into a broader psychological framework. Kornhauser, in discussing the work which had been done before 1944, had this to say:

---

[1] The original report of this study was my doctoral dissertation (13). I am especially grateful to Professors Helen M. Walker and Donald E. Super, respectively member and chairman of my faculty project committee, for their valuable and stimulating guidance during the planning, execution, and writing of the study. Professors Nicholas Hobbs and Albert S. Thompson, who also served on the committee, were sources of helpful criticism and constructive suggestions, the former in the design of the experiment and construction of the questionnaire, the latter in the analysis of the data and interpretation of results. Dr. Leonard Small, Mrs. Verna Small, and Dr. Gilbert David provided much help in organizing the manuscript and in improving the presentation of much of the material. My wife has helped by her constant interest and encouragement and by participation in the analysis of the data. A number of other people have generously assisted in other phases of the work. Dr. Randall B. Hamrick of the Community Advisory Service Center in Bridgeport was instrumental in obtaining a number of subjects in that city. Misses Barbara Fenton and Joan Williams, Mrs. Kathryn Feeney, Mrs. Ellen Gotham, Mrs. Ann Groninger, and Mrs. Mary Ann Ripca all assisted in the statistical computations. To all of these, as well as to all those in the cooperating institutions who distributed the questionnaires, and to the subjects who volunteered their time and effort, I am deeply grateful.

> Generalizations about the importance or unimportance of wages, or advancement, or job security, or personal treatment are shallow and misleading except as they reflect an appreciation of the social-historical changes affecting industry and working people . . . valid conclusions in this matter are like complex clinical judgments, *aided* by whatever scientific evidence is turned up, but necessarily reaching far beyond such data. In the forming of these complex judgments, conceptual guideposts are sorely needed—psychological formulations about motivation and adjustments which will serve to organize thought and inquiry. The development and improvement of these interpretive constructions . . . may prove to be the most valuable contribution psychologists can make in this field (7, p. 143).

Since these words were written there have been many efforts to relate research findings and design to a theoretical framework. The work of Friend and Haggard (2), Roe (10), and Lewin and his co-workers (8), each employing quite different approaches, are notable examples. There have been, of course, many others. Nevertheless, a theoretical framework providing both an understanding of and potency to deal with this general problem is still lacking.

In developing the theoretical conceptualization with which this study is concerned, we have chosen to deal with human needs and their satisfactions. Work is conceived of as potentially satisfying both primary and secondary needs. This theory has been discussed by Hendricks who states: ". . . work is not primarily motivated by sexual need or associated aggression, but by the need for efficient use of the muscular and intellectual tools, regardless of what secondary needs—self-preservative, aggressive, or sexual—a work performance may also satisfy" (3, p. 311). Hendricks expresses dissatisfaction with the limited view with which the psychoanalytic school regards work, stating that "Clinical psychoanalysis . . . has by no means always recognized work pleasure as a primary psychological motivation" (3, p. 316). We believe with Hendricks that work satisfaction consists of both primary and sublimative satisfactions. In this study an attempt is made to examine need satisfaction in work, and its relationship to job adjustment or "satisfaction."

II. A THEORY OF JOB SATISFACTION: OUTLINE OF THE STUDY

A. Development of the Theory

The theory of job satisfaction selected for investigation is not unique in the sense of applying only to satisfaction in work. Work is seen as simply a special area of human behavior and it is felt that whatever psychological mechanisms operate to make people "satisfied" or "dissatisfied" in general also make them satisfied or dissatisfied in their work. States of tension or dissatisfaction[2] are aroused when an individual is unable to satisfy certain of his needs. For any individual in any given situation the amount of tension or dissatisfaction generated is determined by (*a*) the strength of his needs or drives, and (*b*) the extent to which he can perceive and utilize opportunities in the situation for the satisfaction of those needs. It is suggested that to understand why a person is dissatisfied with his job one would have to know the extent to which any of his needs are not being satisfied, and the relative strength of those needs. This theory may be formally stated in the following manner:

---

[2] The term "dissatisfaction" is used to mean a conscious recognition of a state of tension.

> Over-all job satisfaction will vary directly with the extent to which those needs of an individual which can be satisfied in a job are actually satisfied; the stronger the need, the more closely will job satisfaction depend on its fulfillment.

B. Selection of a Method

Utilization of the theoretical formulation in understanding and predicting job satisfaction requires the means for measuring need strengths of individuals and need-satisfying opportunities in various work environments. In this study one technique of need measurement was developed and investigated to determine how well job satisfaction could, in fact, be estimated from a measurement of need strengths and need satisfactions in work. The validation of such a technique would of course provide a most useful tool for both vocational counselors and those engaged in personnel selection and placement.

If such a technique were to be validated there would be an additional step necessary before it could actually be made to serve; namely, the exploration of need-satisfying opportunities in various occupational environments.

This whole procedure is similar to that of developing occupational keys for interest questionnaires. E. K. Strong, in discussing the nature of interests, states, "Expressed liking and disliking . . . indicate that once upon a time under certain circumstances that activity brought satisfaction or dissatisfaction" (14). It should be added that when the response is to an activity in which the person has not engaged, it is in terms of a guess as to whether or not it would be satisfying. Interest measurement, then, is seen as being a kind of need measurement. Norms can be established for predicting both achievement and satisfaction in many occupational groups, and such attempts have met with considerable success. This is possible only because there is a certain communality of needs among the members of one occupational group, and their occupation allows them to derive common satisfactions therefrom.

Since interest measurement has been a fruitful and by no means an exhausted area of research in this field why do we find it necessary, or even desirable, to develop a different technique? Primarily because an interest score is a statistical measure of the likeness between an individual's pattern of responses and the pattern common among and unique to people in a given occupation, and therefore it is most difficult to conceptualize the individual meaning which may be attached to any score, except by examining the responses to the many interest items. A technique which would measure certain specific needs would permit the same kind of empirical analysis and utility as do the interest measures, but would also make possible a more explicit individual understanding. In counseling, to tell a client that he has a high engineering score is useful, but to be able to explore with him his needs and strivings, to relate his personality to the offerings of an occupation, seems much more profitable. The most impelling reasons for developing a new technique were to make it possible to obtain need-strength measures which could be more easily correlated with other psychological data, and to provide the basis for explicit individual understanding and interpretation on the part of research workers and other users of the technique.

There are many possible techniques which might be used in the measurement of need strengths. They can be thought of as lying along a continuum, their position being determined by the structuring of the technique and the influence of consciously determined responses. At one extreme would be free association and the projective devices, the Rorschach and the Thematic Apperception Test. These techniques tap the unconscious needs and strivings of the individual and the material elicited is not readily determined by conscious thought processes. At the other extreme one might list and define a number of needs or need areas and have the respondent rank or rate each of them in terms of its importance to himself. Here the individual would respond in terms of his consciously perceived needs. As we move along this continuum from the unstructured to the structured, the data we elicit become more and more superficial. As we move in the opposite direction, the time, expense, and necessary skill of the administrator all become greater. One of the goals of the study was the development of a technique which could be widely used without requiring extensive, specialized training on the part of the administrator. A fairly well-structured questionnaire was considered to be the most easily administered device. It was thought possible to develop a questionnaire of sufficient subtlety to get beyond what the respondent was willing to report of what he was willing to perceive, and to avoid the effects of cultural value norms upon the expression of certain types of needs.

C. THE NEEDS MEASURED

A structured measurement limits the areas which may be measured. Deciding which needs to measure presented difficulties. Murray's list (9) was a starting point, but some of the needs which he describes could not be measured by a technique as structured as a questionnaire. Also, his list was considered too lengthy for complete inclusion in this initial effort. A list of twelve needs was evolved, most of which are similar to Murray's, although they are defined and named somewhat differently here.

Six criteria were used in selecting the needs. For inclusion each need should be one which:

1. if present in a person is likely to be a rather permanent and stable part of his basic personality structure—an ever-present factor in the determination of adjustment;
2. is judged to be relatively important in the determination of adjustment, although differing in degree of importance among individuals;
3. is present in many people;
4. might conceivably be satisfied in a work environment;
5. is definable and unique;
6. is amenable to measurement by a paper-and-pencil questionnaire.

Twelve needs which met the criteria requirements were chosen. The list of needs with the definition of each is presented below:

A. *Recognition and Approbation.* The need to have one's self, one's works, and other things associated with one's self known and approved by others.

B. *Affection and Interpersonal Relationships.* The need to have a feeling of acceptance by and belongingness with other people. The need to have people with whom to form these affective relationships.

C. *Mastery and Achievement.* The need to perform satisfactorily according to one's own standards. The need to perform well in accordance with the self-perception of one's abilities.

D. *Dominance.* The need to have power over and control of others.

E. *Social Welfare.* The need to help others, and to have one's efforts result in benefits to others.

F. *Self-expression.* The need to have one's behavior consistent with one's self-concept.

G. *Socioeconomic Status.* The need to maintain one's self and one's family in accordance with certain group standards with respect to material matters.

H. *Moral Value Scheme.* The need to have one's behavior consistent with some moral code or structure.

I. *Dependence.* The need to be controlled by others. Dislike of responsibility for one's own behavior.

J. *Creativity and Challenge.* The need for meeting new problems requiring initiative and inventiveness, and for producing new and original works.

K. *Economic Security.* The need to feel assured of a continuing income. Unwillingness to "take a chance" in any financial matters.

L. *Independence.* The need to direct one's own behavior rather than to be subject to the direction of others.

III. THE QUESTIONNAIRE

A questionnaire (12) was designed to measure three variables:[3] (*a*) the strength of each of the twelve needs selected; (*b*) the degree to which each of the twelve needs was being satisfied in the individual's job; and (*c*) the individual's over-all job satisfaction. Since the measurement of the strength of the needs was the most important as well as the most difficult of the three, the largest part of the questionnaire was devoted to it. A five-part questionnaire was developed, of which three parts were devoted to the measurement of need strengths, one to the measurement of need satisfactions in work, and one to over-all job satisfaction. These parts of the questionnaire will be discussed below.

A. MEASUREMENT OF THE STRENGTH OF NEEDS

Sections A, C, and E of the questionnaire, each containing a unique type of item, are devoted to the measurement of need strengths. In designing these items maximum subtlety was a prime consideration.

The items in part A consist of quotations, each representing verbal expression of one of the needs in a "conversational" manner. The respondent is instructed to rate each item on a five-point scale to show how closely the feelings expressed in the quotation match his own feelings. There are 48 items in part A; four items are keyed to each of the 12 need-strength scales.

In part C the items are designed to get at the strength of needs through identifications, value judgments, and projections. These items are also rated to indicate degree of agreement with the idea expressed. They are in multiple-choice format, but differ from multiple-choice items in that every response is rated. There are 36 such items in part C; three are keyed to each of the need-strength scales.

Part E consists of items which ask the respondent to imagine himself in a number of occupations. For each occupation several possible sources of satisfaction are described. The respondent indicates by a rating his estimate of the extent to which each of these described satisfactions would be a source of satisfaction to him were he in that occupation. For these items the rating is again made on a five-point scale. There are 48 such items in part E; four are keyed to each of the need-strength scales.

To recapitulate, the strength of each need is measured by eleven items, four in part A, three in part C, and four in part E, making a total of 132 items devoted to measurement of need strengths.

---

[3] The complete questionnaire is presented in the Appendix.

B. MEASUREMENT OF NEED SATISFACTION IN WORK

Part D of the questionnaire is designed to measure the extent to which each of the 12 needs is satisfied in the respondent's job. The degree of subtlety required in measuring this variable is considered to be less than in measuring the need strengths. Individuals can achieve greater accuracy in perceiving the extent to which a given need is satisfied than in estimating the degree of importance which they actually attach to it. Therefore these items are rather straightforward. For each need, two statements (each somewhat differently worded) were presented. Each expresses a high degree of satisfaction with the given need. The respondent rates his satisfaction on a five-point scale: 5 if he agrees that the need is completely satisfied in his job, 1 if he feels the need not at all satisfied.

C. MEASUREMENT OF OVER-ALL JOB SATISFACTION

Over-all job satisfaction was measured in part B of the questionnaire. There are three items, two of which are similar to those originally used by Hoppock (5, pp. 250-251) in his study on job satisfaction. A third, open-ended question followed these two. In the actual analysis of the data gathered in the study only one of the three items was used as the index of job satisfaction.[4]

IV. THE SAMPLE

The desire for a careful sampling to use in studying the questionnaire and in exploring the validity of the hypothesis was tempered by a number of realistic considerations. The length of the questionnaire and the time demanded for completing it (at least an hour in most cases) made it difficult to obtain subjects. Employed men were sought in a number of business institutions and vocational guidance agencies. They were obtained through the cooperation of personnel executives or other administrators ("contact persons"). The results of the study were in general of little interest to these people because of the somewhat technical and theoretical nature of the investigation, and those who were willing to cooperate did so just to be helpful to the author. In no case was the questionnaire filled out on "company time"; the contact persons had to obtain the cooperation of a number of people in their institutions without offering any reward for the effort.

In no case was it possible for the author either to select the subjects or to take part in the actual distribution of the questionnaires. This was done by the contact persons.

In each of the cooperating business organizations the questionnaire distribution was handled by either a personnel executive or some other high level administrator. The purpose of the study was explained to the contact person and an effort was made to convince him of the necessity for avoiding any personal bias in his selection of subjects. To each questionnaire was attached a stamped envelope addressed to the author at Columbia University. After receiving the questionnaires the subjects were then free to do as they pleased with them without pressure from those who had distributed the questionnaire. They could fill them out and mail them to the

---

[4] The reasons for elimination of the other two items are discussed in Chapter V.

TABLE 1

TYPES OF INSTITUTION REPRESENTED IN THE SAMPLE

Institution	Number of Subjects
Two large industrial manufacturing plants	31
Large department store	13
Two government agencies engaged in technical work	20
Two philanthropically supported vocational guidance agencies	8
Total	72

author or not—there was no way for anyone to know who had and who had not filled out the questionnaires. This guarantee of anonymity also mitigated any feeling of having to give "favorable" answers.

In the vocational guidance agencies the distribution was handled by individual counselors who gave the questionnaires to employed clients and received them when the clients had filled them out. There was no anonymity for the vocational guidance subjects, but it is doubtful if this had any effect on their responses since the questionnaires were administered as part of the psychometric routine.

One hundred thirteen questionnaires were distributed during the spring of 1950, and 72 were returned to the author. Table 1 presents a frequency distribution of the subjects in terms of the type of institution from which they were obtained. Sixty-seven of the respondents filled out a personal data form which enables us to present the age and salary distributions as well as an occupational grouping of the sample. The occupational classification is based on Part II of the *Dictionary of Occupational Titles*. These data are presented in Tables 2, 3, and 4.

TABLE 2

AGE OF THE RESPONDENTS

Age (In Years)	Frequency
15–19	1
20–24	9
25–29	19
30–34	11
35–39	11
40–44	4
45–49	5
50 or over	6
(No response)	(1)
Total	67

TABLE 3

SALARY OF THE RESPONDENTS

Salary (Dollars per Week)	Frequency
$25–29	2
30–39	3
40–49	9
50–59	4
60–69	13
70–79	8
80–89	6
90–99	8
100–109	6
110 or over	5
(No response)	(3)
Total	67

TABLE 4

OCCUPATIONAL CLASSIFICATION OF THE RESPONDENTS

Occupational Classification	Frequency
0. Professional and Managerial	(37)
Professional	22
Semiprofessional	12
Managerial and official	3
1. Clerical and Sales	(20)
Clerical and kindred	10
Sales and kindred	10
4. Skilled	(8)
Skilled occupations in manufacturing	1
Skilled occupations in non-manufacturing	1
Foremen	6
6. Semiskilled	(2)
Semiskilled occupations in manufacturing	2
Total	67

This sample is obviously atypical of the total working force and is certainly not to be considered a random sample of that population. The group is young—very few subjects are over forty. The group is a high level one, with a large concentration of professionals, semiprofessionals, and clerical and sales people, very few semiskilled people, and no representation from the unskilled, service, or agricultural occupations. This selectivity was intentional rather than fortuitous. The first form of the questionnaire is rather long and the language somewhat difficult. It was considered desirable to exclude individuals who might have insufficient verbal facility. Before the questionnaire could be used with subjects having lower verbal ability it would have to be revised and simplified.

The nature of the sample does not destroy the utility of the findings. Although one would not generalize results obtained from this sample to the total employed population without further verification, the psychological factors under investigation here appear universal and not limited to any one occupational level or group.

V. RESULTS

A. STRENGTH OF NEEDS

1. Determination of a Need-Strength Index

As described in Chapter III, the strength of the various needs was measured in three parts of the questionnaire containing 132 items, each keyed to one or another of the 12 needs; these are actually twelve 11-item scales. Every item was rated by the respondents on a five-point scale ranging from 5 (greatest strength) to 1 (least). By adding the 11 ratings assigned to the items on any one scale a value which we call the "absolute scale score" is obtained. The highest possible score for a given scale is 55 and the minimum 11. Using these absolute scale scores, each individual's needs were rank ordered from 1 (the strongest need) to 12 (the weakest need). For each need area both the mean absolute score and the mean rank for the entire sample group ($N = 72$) were computed. The results of this computation are presented in Table 5. They have been arranged in the table in what might be thought of as a need hierarchy for this sample, with the needs which are strongest heading the list.

TABLE 5

MEAN ABSOLUTE SCORE AND MEAN RANK OF EACH NEED

Need Area	Mean Absolute Score of Need	Mean Rank of Need
J. Creativity and Challenge	44.7	2.8
C. Mastery and Achievement	43.8	3.1
E. Social Welfare	40.4	4.9
H. Moral Value Scheme	40.0	5.1
B. Affection and Interpersonal Relationships	39.5	5.2
F. Self-expression	35.9	6.6
D. Dominance	34.7	7.1
A. Recognition and Approbation	34.3	7.5
K. Economic Security	32.8	7.8
L. Independence	30.8	8.6
G. Socioeconomic Status	30.8	8.9
I. Dependence	25.1	10.8

As would be expected, the absolute score values and the rank values each yield identical hierarchies for the whole group. Despite this agreement between the two types of scores for the *group* there is considerable difference between using absolute scores and ranks as indices of need strength for *individuals*. The absolute scale scores are determined by two factors: (*a*) the strength of each need as compared with the individual's other needs and (*b*) error. It is our belief that a large part of this error is due to the systematic differences among individuals in their tendency to use high or low or middle-of-the-

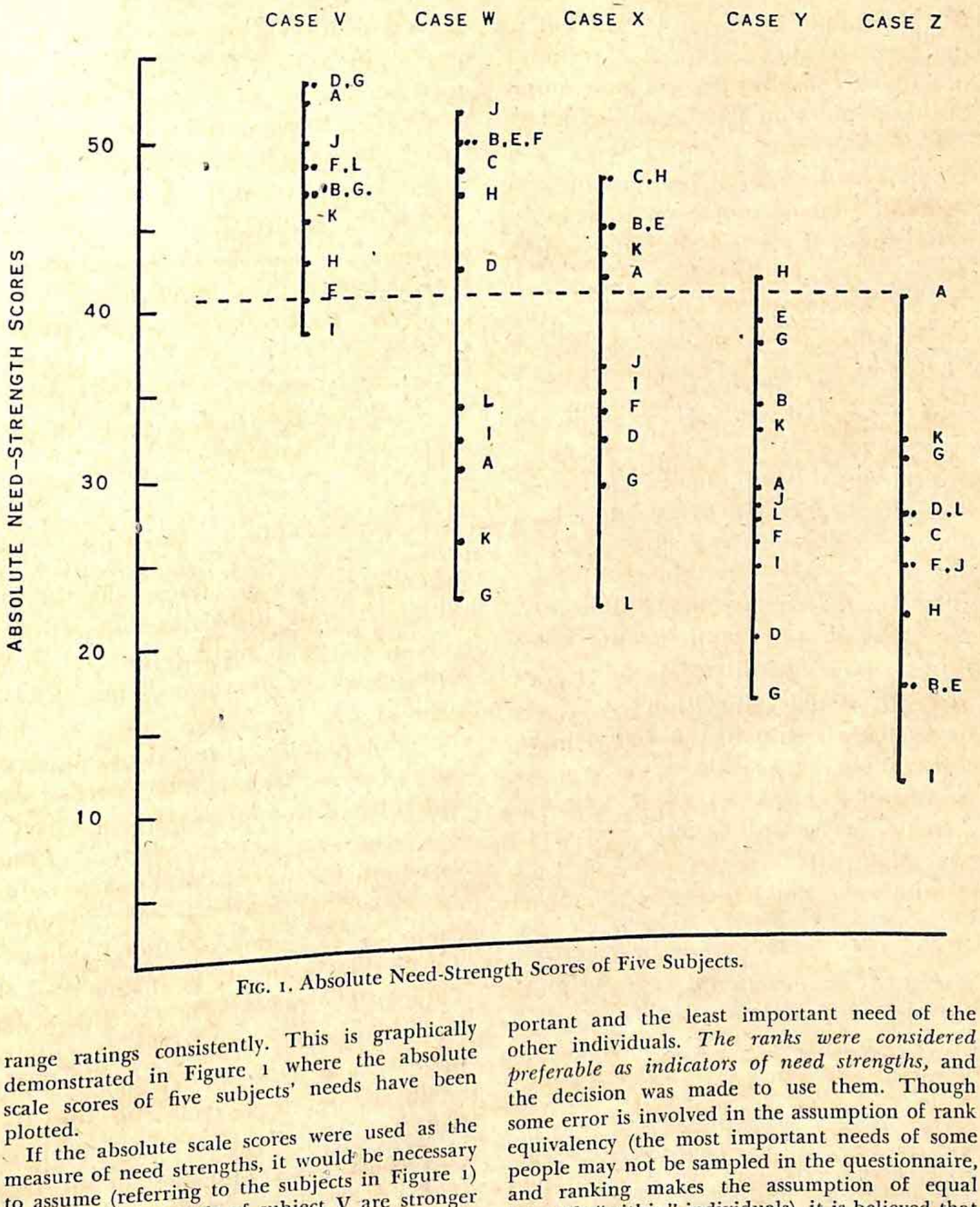


FIG. 1. Absolute Need-Strength Scores of Five Subjects.

range ratings consistently. This is graphically demonstrated in Figure 1 where the absolute scale scores of five subjects' needs have been plotted.

If the absolute scale scores were used as the measure of need strengths, it would be necessary to assume (referring to the subjects in Figure 1) that ten of the needs of subject V are stronger than any of the needs of subject Z. Thus, if the absolute scores were used as indicators of need strength, a need which ranks almost at the bottom of V's hierarchy would be considered as important to him as Z's highest ranking need, since they both have the same absolute score. This is not a very tenable assumption. On the other hand, it is not unreasonable to consider the most important and the least important need of each person equivalent to the most important and the least important need of the other individuals. *The ranks were considered preferable as indicators of need strengths,* and the decision was made to use them. Though some error is involved in the assumption of rank equivalency (the most important needs of some people may not be sampled in the questionnaire, and ranking makes the assumption of equal intervals "within" individuals), it is believed that the error introduced by using rank scores is far less than that introduced by using absolute scores.

In connection with this decision it should be noted that the use of the rank scores also yields a more valid estimate of reliability of the strength scales. This is so because computation of the reliability using the absolute scores would treat that part of the variance produced by the individual rater bias described above (i.e., consistently high, low, or middle-of-the-range ratings) as systematic variance when it is, in fact, error variance. Since this between-persons variance is rather great, the reliability estimates would be spuriously large.

2. *Reliability of Need-Strength Scales*

Because of the analyses which are made later in the paper it is necessary to consider two kinds of reliability with respect to the measurement of need strengths.

The question which the first reliability estimate must answer is: How certain can we be of the rankings of the 12 needs for any individual in our sample? How trustworthy is the hierarchy obtained for each person? This is the question of reliability among the 12 needs "within" individuals (there are 72 such estimates possible—one for each case in the sample). To estimate this reliability 35 of the 72 cases were selected at random. For each case the need strengths were ranked independently on the basis of the absolute scores obtained on part A of the questionnaire, then on the scores obtained on part C, and then part E. These rankings were based on subscales having 4, 3, and 4 items respectively. Coefficients of concordance[5] were computed to indicate agreement among the three sets of ranks "within" each individual. The distribution of these coefficients is presented in Table 6.

TABLE 6

RELIABILITY OF THE NEED-STRENGTH HIERARCHIES FOR 35 SUBJECTS

Coefficients of Concordance among Subsections	Spearman-Brown Approximation of Full-Scale Reliability	% of Cases ($N=35$)	Cumulative %
.70–.74	.88–.90	11	11
.65–.69	.85–.87	9	20
.60–.64	.82–.84	11	31
.55–.59	.79–.81	14	45
.50–.54	.75–.78	16	61
.45–.49	.71–.74	6	67
.40–.44	.67–.70	9	76
.35–.39	.62–.66	9	85
.30–.34	.56–.61	6	91
.25–.29	.50–.55	0	91
.20–.24	.43–.49	6	97
(.06)	(.16)	3	100

This analysis is similar to any split-part reliability estimate. The estimate is based on a number of measurements (two usually, three in this case), each of which is shorter than the one for which the reliability estimate is sought. Since the length of a test or questionnaire affects its reliability, the subpart estimates are too small. The Spearman-Brown prophecy formula provides a correction for the underestimates of the subpart analysis. Since the subparts in this case differ in length and since the rank scores whose reliabilities we seek are not the means of the subpart rank scores, the formula is not strictly applicable in this case and does not provide an exact estimate of the rank-score reliabilities because of these inconsistencies. It is felt, though, that the corrected estimates are considerably closer to the actual values than are the uncorrected estimates based on subparts, each about one-third the length of the total scales. The Spearman-Brown values have therefore been included in the table as being the best possible estimates (the corrections were computed for tripled length). About three-fourths of the full-scale estimated reliability coefficients are .67 or greater.

The other question of reliability of the need-strength scales is this: How certain can we be of the rankings of all the individuals in the sample on any one need? If five individuals were selected, for each of whom the rank of a given need is 1, and five other individuals, for each of whom the rank of that same need is 2, how certain can we be that these two groups of individuals actually differ with respect to the relative strength of that need? This is the question of reliability of each need scale for the entire sample (there are 12 separate estimates required—one for each need-strength scale). Here

---

[5] A coefficient of concordance is equal to the mean rank-order correlation of more than two sets of ranks taken two at a time in all possible combinations. In this case there would be three such combinations: A with C, A with E, and C with E (14).

TABLE 7

NEED-STRENGTH SCALE RELIABILITIES COMPUTED BY ANALYSIS OF VARIANCE

Need Area	Coefficient of Reliability
A. Recognition and Approbation	.53
B. Affection and Interpersonal Relationships	.60
C. Mastery and Achievement	.41
D. Dominance	.64
E. Social Welfare	.92
F. Self-expression	.65
G. Socioeconomic Status	.66
H. Moral Value Scheme	.63
I. Dependence	.44
J. Creativity and Challenge	.73
K. Economic Security	.80
L. Independence	.82

we are concerned with one need and all the cases (even though the reliability of the rank scores on any one scale is obviously dependent on the reliabilities of the other scales); in the previous analysis we were concerned with one case and all the needs. The technique used to estimate these reliabilities is analysis of variance. It is a method somewhat similar to that described by Ebel (1).

The analysis is again based on the rankings on the subparts of the questionnaire. The three subrankings of a given need for any individual are considered to be independent estimates of the actual rank of that need. The total variance of these estimates (based on $3 \times 72 = 216$ estimates) can be separated into: (*a*) variance among individuals, (*b*) variance among subsections, and (*c*) error. The coefficient of reliability for a particular need-strength scale would then be given by this formula:

$$r = 1 - \frac{\text{Mean error variance}}{\text{Mean variance among individuals}}.$$

The results of this analysis are presented in Table 7. Nine of the 12 scales have coefficients of reliability equal to or greater than .60, while two of them are less than .50.

It should again be noted that the reliability estimates will be somewhat inexact because the actual rank scores are not the means of the three subsection ranks, but are arrived at through use of the total absolute scale scores.

In the present study no attempt was made to investigate the reliability of the need-strength scales over a period of time. Obviously, for the instrument to have value in vocational counseling and personnel placement, it is essential that error due to passage of time be slight. Although one of the criteria for selecting the needs was stability, there is no measure of the scales' stability at the present time.

3. Relationships among the Need-Strength Scales

One of the criteria in selecting the needs to be investigated was the independence or uniqueness of each. To determine if this aim was accomplished, coefficients of correlation were computed between the rank scores on each need-strength scale and all the others. The results of this analysis are presented in Table 8.

TABLE 8

INTERCORRELATIONS AMONG THE NEED-STRENGTH SCALES

Variable	A	B	C	D	E	F	G	H	I	J	K	L
A. Recognition and Approbation		−.10	−.23	.25	−.46	−.17	.29	−.34	−.34	.04	−.30	.29
B. Affection and Interpersonal Relationships			.19	−.22	.19	−.20	−.22	−.03	.23	.09	−.20	−.29
C. Mastery and Achievement				−.05	.04	−.12	−.26	−.04	.12	−.01	.00	−.13
D. Dominance					−.43	−.29	.35	−.36	−.23	.14	−.27	.11
E. Social Welfare						−.04	−.56	.33	.40	−.28	.33	−.58
F. Self-expression							−.12	−.09	−.17	.02	−.23	.28
G. Socioeconomic Status								−.34	−.34	.03	−.13	.12
H. Moral Value Scheme									.12	−.09	.06	−.39
I. Dependence										−.29	.29	−.40
J. Creativity and Challenge											−.44	.01
K. Economic Security												−.26
L. Independence												

TABLES 9a and 9b

INTERCORRELATIONS WITHIN TWO CLUSTERS OF NEED-STRENGTH SCALES

Variable	A	D	G	L
A. Recognition Approbation		.25	.29	.29
D. Dominance			.35	.11
G. Socioeconomic Status				.12
L. Independence				

Variable	B	E	H	I
B. Affection and Relationships		.19	−.03	.23
E. Social Welfare			.33	.40
H. Moral Value Scheme				.12
I. Dependence				

An inspectional analysis of the table of intercorrelations indicates that in general the scales are fairly independent of each other. However, it is possible (by inspection) to extract two clusters of scales which are (with one minor exception) positively correlated among themselves but negatively correlated with each other. This is shown in Tables 9a, 9b, and 9c.

TABLE 9c

CORRELATIONS BETWEEN EACH NEED-STRENGTH SCALE IN ONE CLUSTER AND THOSE OF THE OTHER CLUSTER

Variable	A	D	G	L
B. Affection and Relationships	−.10	−.21	−.22	−.29
E. Social Welfare	−.46	−.43	−.56	−.58
H. Moral Value Scheme	−.34	−.36	−.34	−.39
I. Dependence	−.34	−.23	−.34	−.40

In one of these clusters are grouped the needs for affection, for helping others, for living according to some restraining code of behavior, and for being dependent. These needs have in common the deference of the ego—the restraining of assertive or aggressive impulses. Psychoanalytically, the expression of such needs might be interpreted as defenses against the recognition of aggression or hostility.

In the other cluster are the needs for recognition, dominance, status, and independence: ego-assertive needs. Theoretically it is conceivable that the better-adjusted respondents were able to recognize and accept the presence of the assertive needs more easily than the less well-adjusted respondents. This would account for the emergence of these two clusters. Despite these within-sample differences it should be noted that, as might be expected, the ego-restraining and ego-deferential needs are rated stronger for the sample as a whole than are the assertive, possibly aggressive needs (with the exception of the need for Dependence, which is more or less taboo for the adult male in our culture).

We shall refer to these findings later in interpreting the results of the hypothesis-testing analyses.

B. SATISFACTION OF NEEDS

As described in Chapter III, the degree to which each of the needs is being satisfied in the jobs of the respondents was measured in part D of the questionnaire. There are 24 items in this section, two for each need area. The items were rated by the respondents on a five-point scale, 5 representing maximum satisfaction and 1 minimum or no satisfaction. The mean of the two responses was used as the index of need satisfaction (with the exception of three scales discussed below). In order to estimate the reliability of these two-item scales, a product-moment correlation was computed between ratings on the two items in each pair. To arrive at an estimate of the reliability of the mean of the pairs, the Spearman-Brown prophecy formula was used. These

TABLE 10

RELIABILITIES OF THE NEED-SATISFACTION SCALES

Need Area	Correlation between Item Pairs	Estimated Full-Scale Reliability (Spearman-Brown)
A. Recognition and Approbation	.66	.80
B. Affection and Interpersonal Relationships	.20	.33
C. Mastery and Achievement	.29	.45
D. Dominance	.61	.76
E. Social Welfare	.49	.66
F. Self-expression	.06	.11
G. Socioeconomic Status	.63	.77
H. Moral Value Scheme	.48	.65
I. Dependence	.71	.83
J. Creativity and Challenge	.72	.84
K. Economic Security	.78	.88
L. Independence	.53	.70

values are presented in Table 10. Nine of the scales have reliability coefficients of .65 or greater.

Since these scores were to be used in a correlation analysis, it was felt that the reliabilities of scales B (Affection and Interpersonal Relationships, $r = .33$), C (Mastery and Achievement, $r = .45$), and F (Self-expression, $r = .11$) were so low as to preclude any useful analyses. Examining the low-reliability pairs with hindsight's sharpened perceptions and comparing them with the more reliable pairs (see Appendix) leaves no cause for wonder at the low coefficients obtained. Two possible courses of action are: (*a*) not to use the three unreliable scales in any of the further analyses, or (*b*) use one item from each pair for the need-satisfaction score. Since there are only 12 needs, the first alternative seems unnecessarily wasteful. In each of the three pairs it is possible to choose the more desirable item with little difficulty. In the Affection and Interpersonal Relationships scale, item D2 relates directly to satisfaction derived on the job, whereas item D16 refers to off-the-job satisfactions and is in terms of time rather than the need itself. In the Mastery and Achievement scale, item D3 seems to relate directly to the concept involved, whereas item D14 does not and is, moreover, somewhat confusing since it asks two questions rather than one. In the Self-expression scale, it was felt that the interpretation of "self-expression" in item D6 would vary from subject to subject. This expression was not used in measuring the strength of that need; instead, the concept of being one's self or "letting yourself go" was used. Therefore item D2 was chosen as the measure of satisfaction of the Affection and Interpersonal Relationships need, D3 for the Mastery and Achievement scale, and D21 for the Self-expression scale. That satisfaction of each of these needs is being measured by only one item should be borne in mind.

It should be noted that the reliability estimates of the need-satisfaction scales, like those of the need-strength scales, do not take into account any fluctuation with the passage of time. Stability over periods of elapsed time is as necessary in the case of need satisfactions as it is for need strengths for the technique to be of practical use.

Using the scores described above, for each need the mean satisfaction score and the standard deviation of the satisfaction scores were computed for the sample. These are presented in Table 11.

To make Table 11 more meaningful, the scale values which the respondents used are described here:

(5) completely satisfied
(4) very well satisfied
(3) well satisfied
(2) slightly satisfied
(1) not at all satisfied

TABLE 11

MEAN AND STANDARD DEVIATION OF NEED-SATISFACTION SCALES ($N = 72$)

Need Area	Mean Satisfaction Score	*SD*
H. Moral Value Scheme	3.94	1.16
J. Creativity and Challenge	3.70	1.16
F. Self-expression	3.68	1.10
B. Affection and Interpersonal Relationships	3.67	1.09
L. Independence	3.60	.96
I. Dependence	3.58	1.28
K. Economic Security	3.43	1.22
E. Social Welfare	3.40	1.06
C. Mastery and Achievement	3.20	1.39
A. Recognition and Approbation	3.00	1.12
G. Socioeconomic Status	2.83	1.06
D. Dominance	2.73	1.19

In general there is no uniformity of feelings of satisfaction in any of the need areas. The standard deviations of about 1 indicate that in each area there are some subjects who are well satisfied and some who are not, since the mean values tend to center around the neutral or slightly favorable point. (If each of the 5 ratings on any scale occurred with equal frequency, the standard deviation would be 1.41.) It is interesting to note that the needs which we have called the "passive, hostility-restraining" needs are in general felt to be more satisfied than the assertive, aggressive needs. Is it possible that our hypothesized less well-adjusted people not only must feel that these needs are important to them, but must feel that they are being satisfied?

Because these scales are later used in correlation analyses, the fact that the standard deviations are fairly homogeneous should be noted. Were they markedly varied, interpretation of the correlation coefficients would be made more difficult.

C. Over-all Job Satisfaction

There were three items on part B of the questionnaire designed to measure over-all job satisfaction. One of these items is an open-ended question; it was intended to have the responses to this item rated by several judges in terms of over-all satisfaction. Unfortunately the majority of the responses did not lend themselves to this kind of judgment. The other two items are similar to those developed by Hoppock (5, pp. 250-251). A number of people who examined the questionnaire (other than the respondents) felt that item B2 (see appendix) was confusing and difficult to answer. Since the third item (B1) seemed to get at over-all satisfaction in a direct and simple way, it was used in the analyses.

Item B1 is presented below with a frequency distribution of responses to it.

Item B1:

If I were to rate my general, over-all feelings about my job, I would say that I *usually:*

a. like it *extremely well*	16
b. like it *very much*	31
c. like it *quite a bit*	11
d. like it *fairly well*	9
e. like it *just a little*	0
f. neither like nor dislike it	3
g. *dislike* it	2
h. *hate* it	0

In assigning values to this scale it was recognized that the intervals were probably not equivalent. No other valid assumption about the size of the intervals can be made however, and, as was done in the need-satisfaction scales, linear values were assigned, $a = 8$, $b = 7$, and so on.

The distribution of responses to item B1, given above, is similar to others which have been obtained in samples of employed people. Usually there is a highly skewed distribution with very few people expressing strong dissatisfaction with their jobs. We cannot attempt an absolute valuation of the responses to this item and decide who are the "satisfied" and who the "dissatisfied." Our only assumption is that those who answer higher on the scale are taken to be the more satisfied, those at the lower end the more dissatisfied.

In treating the need-strength scores a correction was made for what we referred to as respondent "bias"—a tendency to use responses at one part of the scale—and ranks were used rather than absolute scores. This was not done for the need-satisfaction scales. It was not difficult to conceive of some people having most of their needs well satisfied in their work and others not. We could, therefore, accept the possibility that one subject's need *satisfactions* are all greater than

TABLE 12

FREQUENCY DISTRIBUTIONS OF (1) THE SUMMED NEED-STRENGTH RATINGS, (2) THE SUMMED NEED-SATISFACTION RATINGS, AND (3) THE OVER-ALL JOB-SATISFACTION RATING FOR THE 20 SUBJECTS HAVING THE HIGHEST SUMMED NEED-STRENGTH RATINGS (H) AND THE 20 HAVING THE LOWEST (L)

Sum of Need-Strength Ratings	Frequency Distributions H	L
560–579	1	
540–559	2	
520–539	1	
500–519	3	
480–499	9	
460–479	4	
440–459		
420–439		
400–419		
380–399		7
360–379		5
340–359		6
320–339		1
300–319		0
280–299		1
Total	20	20

Sum of Need-Satisfaction Ratings	Frequency Distributions H	L
114–117	1	0
110–113	1	1
106–109	3	0
102–105	1	0
98–101	0	1
94– 97	1	1
90– 93	2	2
86– 89	2	1
82– 85	3	1
78– 81	1	3
74– 77	3	4
70– 73	0	1
66– 69	1	2
62– 65	0	3
58– 61	0	0
54– 57	1	0
	20	20

Over-All Job-Satisfaction Ratings	Frequency Distributions H	L
8	6	3
7	7	10
6	3	4
5	3	2
4	0	0
3	1	1
	20	20

another's, although in the case of need *strengths* the assumption of rank-equivalence was considered desirable as a correction for rater bias. In the case of the over-all satisfaction scores there was no choice but to use the absolute values.

A valid question then is whether the respondent bias on the strength-of-need scales is carried over to the other scales. Do the people who tend to use high (or low) ratings on the need-strength scales (parts A, C, and E of the questionnaire) *also* use high (or low) ratings in the need-satisfaction and over-all job-satisfaction scales (parts B and D of the questionnaire)? If this is the case, an error has been committed in the treatment of the data. To see if this does occur, the 20 individuals whose summed need-strength ratings are highest were compared to the 20 whose summed need-strength ratings are lowest. Table 12 allows a comparison of the distributions of these two groups in terms of (*a*) the sum of each person's need-strength ratings (for the 132 items in parts A, C, and E of the questionnaire), (*b*) the sum of each person's need-satisfaction ratings (for all the items in part D of the questionnaire), and (*c*) each person's over-all job-satisfaction rating (item B1).

Although the distributions of the two groups are markedly different with respect to need-strength ratings (as might be expected—see Figure 1), they do not differ significantly with respect to their need-satisfaction or over-all satisfaction ratings. We can conclude that the *tendency to use high or low ratings in responding to the need-strength scale items is not carried over to the other two scales.*

D. RELATIONSHIP BETWEEN NEED STRENGTH AND NEED SATISFACTION

A possible relationship between need strengths and need satisfactions was suggested to the author. Perhaps the degree of frustration of a need affects the perceived strength of that need. A need which is well satisfied may not be felt to be as strong as one which is frustrated. Were this true, it would cast some doubt on the validity of the need-strength measures. Two analyses were made to test this possibility.

Thirty-five cases were selected at random from the sample. The need satisfactions for each were ranked, 1 being the most satisfied need and 12 the least satisfied. Then rank-order coefficients of correlation were computed between the need strengths and the need satisfactions for each individual. Were the hypothesis true, these coefficients would be negative in sign. A frequency distribution of these coefficients is presented in Table 13. They vary in size from .71 to —.45; the median value is .08. This distribution

TABLE 13

RANK-ORDER CORRELATION COEFFICIENTS BETWEEN RELATIVE NEED STRENGTHS AND RELATIVE NEED SATISFACTIONS ($N=35$)

Rank-Order r	Frequency
.61–.71	2
.50–.60	4
.39–.49	3
.28–.38	3
.17–.27	2
.06–.16	4
−.05–.05	7
−.16–−.06	2
−.27–−.17	4
−.38–−.28	3
−.49–−.39	1
Total	35

does not lend support to the suggested hypothesis.

In a second test of the hypothesis the 30 subjects having the highest need-strength scores on each need were compared with the 30 having the lowest. Were the hypothesis true, those scoring higher on a particular need-strength scale would be expected to have a lower satisfaction score. The results of this analysis showed that, for 6 of the 12 needs, the difference is in the hypothesized direction, and for 6 is in the reverse direction. There is no significant difference in need-satisfaction scores between those having higher and lower need-strength scores.

These two analyses indicate that the hypothesis concerning the relationship between the two variables is not substantiated.

E. JOB SATISFACTION AS RELATED TO NEED SATISFACTION

The purpose of this study was to develop an instrument which could be used successfully to estimate job satisfaction from measures of need satisfaction and need strength. In this section a number of analyses are presented which help to evaluate the utility of the technique.

The first and most simple relationship studied was that between need satisfaction and job satisfaction, without taking need strengths into account. Using the entire sample a product-moment correlation was computed between over-all job satisfaction and the mean of the twelve need-satisfaction scores for each individual. The value of the obtained r was .44.

According to the theoretical formulation upon which this study is based, the extent to which needs of varying strength are satisfied should have differing effects on over-all satisfaction. Therefore, to see the effect of taking into account the differences in strength of the needs, the satisfaction score of each need was multiplied by its rank strength score. The mean of these products for each individual is a weighted need-satisfaction score. Over-all satisfaction was then correlated with this value and the resulting r was .48. The improvement in prediction over the .44 correlation of the unweighted satisfaction scores is insignificant.

Several other analyses were made in an attempt to understand the absence of significant improvement. Correlations were computed between: (*a*) over-all satisfaction and satisfaction of each person's *most* important need, (*b*) over-all satisfaction and the mean satisfaction of each person's *two* most important needs, (*c*) over-all satisfaction and the mean satisfaction of each person's *three* most

TABLE 14

CORRELATION BETWEEN OVER-ALL JOB SATISFACTION AND THE SATISFACTION OF SUCCESSIVE NUMBERS OF NEEDS FROM ONE TO TWELVE

Successive Numbers of Needs	Correlation between Mean Satisfaction of These Needs and Over-all Satisfaction
a. first need	.54
b. first 2 needs	.58
c. first 3 needs	.57
d. first 4 needs	.51
e. first 5 needs	.49
f. first 6 needs	.47
g. first 7 needs	.44
h. first 8 needs	.44
i. first 9 needs	.45
j. first 10 needs	.45
k. first 11 needs	.46
l. all 12 needs	.44

important needs, etc. The results of this analysis are presented in Table 14.

The results in Table 14 indicate that, with equal weightings, the optimum prediction can be made by using the first two or three needs and their satisfactions. If the reliability of the scales could be increased, the prediction would probably improve with the inclusion of one or two more needs. The use of multiple correlation might also be expected to improve the prediction. There seems to be a fairly good indication here that the measure of the extent to which each person's *most important* needs are satisfied yields the best prediction of his over-all satisfaction when equal weightings are used.

The question of whether *any* two or three needs would provide the same results might be explored. Suppose one were to study the relationship between the satisfaction of *any* two needs of each person and over-all satisfaction, might it not prove to be as great as for those measured as strongest? To test this and to explore further the data in the light of the theory, the following analysis was made.

TABLE 15

CORRELATIONS BETWEEN OVER-ALL JOB SATISFACTION AND THE SATISFACTION OF NEEDS OF DIFFERING STRENGTH

Strength Rank of Need for Individual	Correlation between Over-all Satisfaction and Degree of Satisfaction of Need in *n*th Rank
1	.54**
2	.47**
3	.36*
4	.18
5	.21
6	.10
7	.10
8	.20
9	.34*
10	.20
11	.30*
12	.13

** r is significantly different from zero at .001 level.

* r is significantly different from zero at .01 level.

TABLE 16

CORRELATION BETWEEN OVER-ALL JOB SATISTION AND SATISFACTION IN EACH OF THE NEED AREAS

Need Area	Mean Rank	Correlation between Over-all Satisfaction and Satisfaction of Given Need
J. Creativity and Challenge	2.8	.51**
C. Mastery and Achievement	3.1	.42**
E. Social Welfare	4.9	.47**
H. Moral Value Scheme	5.1	.20
B. Affection and Interpersonal Relationships	5.2	.12
F. Self-expression	6.6	.15
D. Dominance	7.1	.27
A. Recognition and Approbation	7.5	.32*
K. Economic Security	7.8	.03
L. Independence	8.6	.02
G. Socioeconomic Status	8.9	.40**
I. Dependence	10.8	.14

** r is significantly different from zero at .001 level.

* r is significantly different from zero at .01 level.

We have already obtained the correlation between the satisfaction of each person's most important need and over-all satisfaction. Next, the same relationship was measured for each person's second strongest need alone; then for the third, and for each of the others. These correlations are presented in Table 15. To be consistent with the theory, the first correlation should be highest, and the others smaller in descending order. As expected, the first is highest, and the next few descend in order, but there are several unexpected high correlations at the lower end of the table. Is this psychologically explicable, or a chance occurrence, or is it due to some peculiarity of the instrument or the techniques of analysis employed? On suspicion that the latter was responsible, the data were examined in one more fashion. The mean ranking for the whole sample on each of the needs in Table 5 shows a considerable homogeneity among the individuals in the sample in the rankings of the relative strengths of the needs. Were this not so, the mean ranks, instead of running from 2.6 for Creativity and Challenge to 10.8 for Dependence, would all tend to center around 6, 7, and 8. We can, therefore, conceive of this listing of needs as forming a hierarchy for the sample. Theoretically, correlating satisfactions with each of the needs in the hierarchy and over-all satisfaction should result in a series of correlations, the largest one occurring with the Creativity and Challenge scale, the strongest-felt need for the sample, and the smallest with the Dependence scale, the weakest-felt. The correlations should array themselves in descending order. They are shown in Table 16.

There is a marked similarity between Tables 15 and 16. However, in Table 16 one can see what caused the unexpected high correlations in Table 15 to occur. Three needs seem out of place in Table 16 (assuming the validity of the hypothesis). They are Dominance, Recognition and Approbation, and Socioeconomic Status, three of the four assertive, aggressive needs. What has probably occurred is that a true measure of their strengths was not obtained—they are more important than people will admit or, more likely, than they are able to perceive. Our efforts at subtlety in the construction of the need-strength measures does not seem to have been successful in the case of these three needs at least.

In summary, the most accurate prediction of over-all job satisfaction can be made from the measure of the extent to which each person's strongest two or three needs are satisfied. Since there is considerable homogeneity in the sample with respect to the relative importance of each of the needs measured, it makes little difference whether the individual's or the group's need hierarchy is used as the criterion of relative need strength. We would not expect this to be true in the case of a more heterogeneous group. The accuracy with which over-all job satisfaction may be estimated is limited by the probable existence of a constant error in the measurement of need strengths of three of the assertive or aggressive needs: viz., Dominance, Recognition and Approbation, and Socioeconomic status.

VI. IMPLICATIONS FOR FUTURE RESEARCH

The results described in the previous chapter are, in general, consistent with the theoretical framework upon which the study was based. More important is the expectation that further research and investigation of the method would be rewarding. The major weaknesses in the present instrument are believed to be fairly well defined, thereby pointing rather directly to certain subsequent steps. The most serious question concerns the validity of the need-strength scales. It is accepted here that people are unable to perceive with any great accuracy the strengths of their own needs. Despite this, needs can be measured with some success through the use of projective techniques. The question is whether it will be possible to duplicate the success of the projective techniques through revision of the present instrument and, as was our goal, to provide a useful tool in this area for those untrained in the use of projective techniques. It appears clear that increased subtlety of the need-strength measures is an essential preliminary to improvement of the present method.

As part of the work of improving the accuracy of the need-strength measurement, it is suggested that a validity study be conducted, comparing the results obtained on the questionnaire with other measures: clinical appraisals, projective measures.

As suggested earlier in the paper, study will be necessary to determine if it is possible to develop need-satisfaction profiles (or norms) for broad occupational groups (necessary if the method is to be used in guidance work) and for specific work environments (necessary if the method is to be used in personnel selection and placement).

These, briefly, are the essential steps which must be taken before the method can be fully evaluated and put to use if found adequate.

VII. SUMMARY

This study describes an attempt to develop a theoretical conceptualization of job satisfaction having functional utility. The theory deals with human needs and their satisfaction. It is felt that the mechanisms which operate to make people satisfied or dissatisfied in general also make them satisfied or dissatisfied in their work. The theory formally stated is this:

> Over-all satisfaction will vary directly with the extent to which those needs of an individual which can be satisfied in a job are actually satisfied; the stronger the need, the more closely will job satisfaction depend on its fulfillment.

Twelve needs were selected for investigation and a questionnaire constructed to determine how well job satisfaction could, in fact, be estimated from a measurement of need strengths and need satisfactions in work. The questionnaire contains five parts, three measuring need strengths, one measuring need satisfactions in work, and one measuring over-all job satisfaction (see Appendix).

The questionnaire was administered to a sample of 72 employed men. The subjects were mostly in the professional and semiprofessional occupational groups.

For this sample the strongest needs, as reported by self-ratings, were those for Creativity and Challenge, Mastery and Achievement, and Social Welfare (helping others); the weakest were those for Independence, Socioeconomic Status, and Dependence.

In analyzing the intercorrelations among the need-strength scales, two clusters of needs could be extracted. One group contained needs which were passive or hostility-restraining in nature; the other contained assertive, aggressive needs.

Several analyses were made to determine how well job satisfaction could be estimated from the other measures obtained. A coefficient of correlation equal to .58 was obtained between the mean satisfaction score of each person's two strongest needs and over-all satisfaction. Using the mean satisfaction of the three (or more) strongest needs reduced the accuracy of prediction.

There is some evidence which indicates that several of the assertive, aggressive needs are not being accurately measured, and that people tend to underrate the importance to themselves of these needs.

APPENDIX

OCCUPATIONAL ATTITUDES SURVEY[1]

(1) NEED-STRENGTH SCALES: PARTS A, C, AND E.
(2) NEED-SATISFACTION IN WORK SCALES: PART D.
(3) OVER-ALL JOB SATISFACTION SCALE: PART B.

The letters in the parentheses before each item serve to key it to one of the twelve needs in the list below. (In the questionnaire itself the parentheses were used by the respondents for marking their ratings.)

(A) Recognition and Approbation[2]
(B) Affection and Interpersonal Relationships
(C) Mastery and Achievement
(D) Dominance

---

[1] (12), Copyright 1950, Robert H. Schaffer, New York, N.Y.

[2] For definitions of the needs, see Chapter II, C above.

(E) Social Welfare
(F) Self-expression
(G) Socioeconomic Status
(H) Moral Value Scheme
(I) Dependence
(J) Creativity and Challenge
(K) Economic Security
(L) Independence

DIRECTIONS

closely your feelings match those of the speaker. A rating of (5) would indicate that your feelings naire all at once, but it is best if you answer each part in order and do not skip around.

Please answer every question (except those concerning your identity, which you may omit if you wish).

Do not spend too much time answering each item. Your first impression is the important one.

Read the directions for each part very carefully. Since each part of this questionnaire is different, it is necessary to read the directions for each part to answer the questions correctly.

You are reminded that nobody except the investigator and his staff will ever see these questionnaires once they are returned. Employers and other cooperating groups will see only tabulated group results in which it will be impossible to identify any individual, so please feel free to answer the questions as truthfully and frankly as possible.

PERSONAL DATA

Name: ....

Address: ....

....

....

(You may omit the answers to the above if you wish)

Age: ....

Name of your Occupation: ....

Describe the work you do: ....

....

....

Name of your employer: ....

What business is your employer in? ....

How long have you been in your present *job?* ....

How long have you been in your present *occupation?* ....

What are some of the other occupations you have been in?

Occupation	Length of time
....	....
....	....
....	....

What is your present salary? $................ per week.

Would you be interested in finding out the results of this survey?

yes .... no ....

Would you be willing to be contacted by one of the members of the staff for the purpose of having a one hour interview to discuss your answers to this questionnaire?

yes .... no ....

PART A

DIRECTIONS. In this section you will find statements made by different people when discussing certain aspects of their life and their work. As you read each statement see how closely the *feeling* expressed by the speaker *matches your own feelings. DO NOT* consider the actual situation the person is discussing, but try to understand the attitude he is expressing. Then rate each statement on how closely your feelings match those of the speaker. A rating of *(5)* would indicate that your feelings are in complete agreement with the speaker. A rating of *(1)* would mean that you don't agree at all with the speaker. The numbers between 1 and 5 are used to show amount of agreement. Feel free to use *all the numbers* in this table to show how much your feelings match those of the speaker:

(5) complete agreement
(4) very close agreement
(3) fairly close agreement
(2) slight agreement
(1) no agreement at all

Be sure to rate every statement.

(A) 1. "The nice thing about this job is that when you do something well you know that you're going to *get the credit for it.*"

(B) 2. "When you work as closely as we do it sometimes is difficult to prevent arguments. But I wouldn't trade it for anything. I've been working with this bunch for quite a few years and I think that's about the best thing here—the fact that I have *so many close friends.*"

(C) 3. "Yes, I sometimes work pretty hard even when I don't have to. But that's just so I can know I did a good job. It's important for me to do a *good job according to my own standards.*"

(D) 4. "I like being supervisor. There is a good feeling of satisfaction that comes with being able to give the orders rather than having to take them."

(E) 5. "When I think of the number of people that are benefitted as a direct result of my work I really feel swell. I wouldn't be doing anything else. *Doing work like this that gives me a chance to help those who need it is very important to me.*"

(F) 6. "I like to be myself. That may sound funny, but just think of all the times when you have to act in a way that's different from what you know you are. None of that for me. I always like to act just as I think of myself."

(G) 7. "The reason I moved into this neighborhood is that here I can own one of the nicest houses on the block. If I moved into a higher-class community I would probably be living in a house that was far from being the nicest. I like the place I live in to be at least as nice as most around."

(H) 8. "I get my ideas of right and wrong from the Bible. Others may get theirs from other sources maybe; but just so long as you do have some set of rules by which you live, you're O.K. *It's important to have your life governed by something 'higher' like that.*"

(I) 9. "I refused the chance for promotion for just one reason. Here my work is laid out for me and *I know just what is expected of me.* I'm working under the supervision of one of the best guys in the outfit. I like working for a guy like that—I know I'll get all the help and advice I need."

(J) 10. "It may give you an uncomfortable feeling to have to try something new on your own, something that hasn't been done before, *but I really like it.*"

(K) 11. "What keeps me here at this place is that I know my job is permanent. I may not move ahead very rapidly or get such big salary increases, but I know that this company is here to stay and so is my job. That's what I need to feel comfortable—the knowledge that *my job is always going to be here for me.*"

(L) 12. "Well, it's just that I don't like anybody telling me how to do things. Sure I take advice and orders from my bosses, but I don't particularly like it. *I mostly like to do things on my own,* without having somebody telling me how."

(A) 13. "I like to know that people know what I'm doing and that they appreciate it. I like to get some kind of appreciation from somebody that counts when I do something."

(B) 14. *"I like being with people.* I like my friends where I work; I like meeting new customers, and I like spending lots of time with my wife and kids."

(C) 15. "When I do something I like to do it well! That's a very important thing with me—knowing that I've done a good piece of work."

(D) 16. "I like being foreman because it gets you out of the rut and a little bit on the top of the heap. I like being able to tell my men how to do things correctly, and it gives me a good feeling that I know more than they and can tell them how things are to be done."

(E) 17. *"I get a great kick out of helping people.* I belong to a club that works to help orphans in Europe. I also spend some time working with poor people in this city. There's nothing as self-satisfying as knowing that you've done something to help some person not as well off as you."

(F) 18. "What I like about this job is that I can *say* what I think, *do* what I think I ought to and *act* just the way I feel."

(G) 19. "I like to be able to enjoy the things my friends have. If all my neighbors have cars, I want to be able to have one; if they can send their kids to a summer camp, I want to be able to send mine to one at least as nice."

(H) 20. "I have an idea of the important things in life. I know right from wrong and I live according to those ideals. I don't care what I could get for it, or what it is, if I don't think a thing is right, I just won't do it. For instance, I just junked my old car. If I had sold it to a dealer he would have sold it to somebody else, and the thing just doesn't have any more left to it. I may have lost money on the deal, *but I live by a code."*

(I) 21. "It's really a pleasure not to have the headaches of responsibility. When I have to do something new I always like to get somebody who knows more than I do to show me."

(J) 22. "I like to keep bumping into new problems. I like to figure out new ways of doing things and inventing things. I really enjoy having a problem that takes a lot of figuring on my part—it makes my work really interesting."

(K) 23. "I must know that my future income is secure. I don't need too much to make me happy, but I would like to be *very sure* that I'll have my income no matter how long I live."

(L) 24. "I like to be my *own* boss. Even though this job is not as good as some I might be in, at least here there is nobody to tell me what to do and how to do it. I work my own hours and do pretty much as I please. There's nothing like it."

(A) 25. "I think those people who write anonymous poems must be nuts. What's the sense of doing anything worth-while if you're not going to get the credit for it? That's one of the most important things to me. When I do a good job I want people to know I did it."

(B) 26. "One of the best things in life is having a family that is always with you through thick and thin, and the same with old friends."

(C) 27. "Sure I like a hard job. Now look, I don't mean that I enjoy knocking myself out, but when I do something I want to be able to say when I'm finished, 'Well, George, you did a *darn good job.'* That gives me an awful lot of pleasure."

(D) 28. "There's really something to being a leader. It gives you a terrific feeling if you know more than those around you and if you can tell them how to do things correctly."

(E) 29. "Some folks down at the office call me the 'Boy Scout,' but that doesn't worry me. I just get a big kick out of helping people. If you can help some down-and-outer or give some poor kid a break it really makes your heart feel good. I may not be getting as much for myself as I could, but I'm willing to take the time and trouble to help others because it does me so much good."

(F) 30. "I have different moods like everybody else. What I want is to be able to let myself go and express those moods and feelings. Being able to express myself like that is very important to me."

(G) 31. "Keeping up with the Joneses? Sure I do. Doesn't everybody? I just feel uncomfortable if the people around me have more than I do."

(H) 32. "Living a good life according to my beliefs is very important to me. Some call it religion, some call it morals, or some just plain conscience; but whatever they call it, it means the same thing. I want *all* my actions to jibe with my fundamental beliefs."

(I) 33. "Why push and struggle to get to the top? Whenever I have to do something I haven't done before, I know Stan is always around to show me. It sure is comfortable to know that there is always somebody to help you when you need it."

(J) 34. "You know when I'm at my best? Just give me a tough problem—one where you can't get all the answers out of a book. That's when I can really get into it. Having to use my imagination to dream up some new things or to invent some new methods of solution is something that really gives me a kick."

(K) 35. "To me taking a chance is pretty silly. Why risk what you have for what you *might* get? I keep my money in the bank and don't fool around with investments. I just *couldn't take a chance* on losing anything."

(L) 36. "I don't like to have somebody over me. Even though my boss is nice, I just don't like the idea that he can direct me and tell me what to do. It's the same in my social life; I don't ever hang around with people who will tell me what to do. Even if they know more than I do, I still don't like the idea."

(A) 37. "You know I don't think I would enjoy doing any kind of work no matter what it might involve unless I got *some kind* of recognition for it."

(B) 38. "I just couldn't stand working by myself or being alone for much of the time. I like to have other people around me most of the time whether it's here at work or when I'm off."

(C) 39. "I like to do work that's just at my level. What I mean by that is that I like to do work that's challenging and yet that's easy enough for me to do a decent job at it."

(D) 40. "When they told me that I was going to take Sam's place (the straw boss on the job) I hesitated for a while, but when I thought how good it would be to be *giving* orders for a change instead of *taking* them, I decided that it was just what I needed."

(E) 41. "I guess when I really stop and think about it, my greatest satisfaction in life comes from helping others."

(F) 42. "About the best thing a man can do over the week-end is to let off a little steam. Some guys do it by bowling, others pull weeds out of their garden and others paint pictures. It's all for the same purpose though, and I think it's an important one. All week you have to go around acting like a stuffed shirt, so what better thing could you do on your time off than to act the way you like to—the way you *really* are."

(G) 43. "What counts with me is not just the amount of money I make, but whether my salary is about as large as other people I know. I hate to have to get along without many of the things my friends can afford."

(H) 44. "Look, I was offered a job paying more than this one by the XYZ Company, but I turned it down. The stuff they're making just isn't worth selling, and I don't care how much money I could make out of it, I don't think I'd sell junk like that. It's not worth what I'd have to put up with my conscience."

(I) 45. "If a man isn't really cut out to be great, the best thing he can do is to associate with somebody and to work under somebody who *is*. By working under a really top-notch man one can share a lot of the satisfactions of his accomplishments."

(J) 46. "To me it makes a job more interesting if I have to keep doing *new* things *on my own*, even if I have to take a chance on doing them wrong."

(K) 47. "I just couldn't feel relaxed at all unless I know that my job is pretty permanent and that when I'm unable to work any more that I'll have my retirement pay."

(L) 48. "Well, Burt's job is pretty good, and I guess I could have had it if I wanted, but the trouble is that the boss is around too much there. I like working on my *own* hook most of the time without *anyone* around giving orders."

Be sure you have rated all 48 items in this part before going on to Part B.

PART B

DIRECTIONS. In this part you are to think about your **overall** feelings and attitudes about the job you now have. Try to be as accurate as possible in answering—remember, the information which you give as an individual will be held in strict confidence.

1. If I were to rate my general, overall feelings about my job, I would say that I *usually*

(check *one*)

...... a. like it *extremely well*
...... b. like it *very much*
...... c. like it *quite a bit*
...... d. like it *fairly well*
...... e. like it *just a little*
...... f. neither like nor dislike it
...... g. *dislike* it
...... h. *hate* it

2. If I were to compare my feelings about my job with the feelings of all other working people (or, at least the ones I know) I would say that

(check *one*)

...... a. almost everybody likes their jobs better than I do

...... b. many people like their jobs more than I do

...... c. I like my job about as well as half of other people

...... d. I like my job more than many people

...... e. I like my job better than most people

...... f. I like my job better than practically anybody else

3. If I were to describe my overall satisfaction with my job (that is, how much I *like* it or *dislike* it) in one or two short sentences, I would say:

....

....

....

....

PART C

DIRECTIONS. In this part you will find items relating to your personal *likes* and *dislikes*. Here you are to rate each item on how much you *like* or how closely you *agree with* the idea of each item. Use the numbers from (*1*) to (*5*) to show how much you like or agree with the idea according to this table:

(5) like or agree with idea *completely*
(4) like or agree with idea *very much*
(3) like or agree with idea *quite a bit*
(2) like or agree with idea *slightly*
(1) don't like or agree with idea at all

Rate every item independently, using *any* or *all* of the numbers as often as you like. *Be sure to rate every item* by putting a number in *every* space.

When I see a play or movie, I usually like, or at least feel sympathetic with the individuals who—

(H) 1. do what they think is right regardless of personal consequences; those who never violate their principles of life.

(D) 2. seem to be the leaders; the ones who can tell others what to do and who are respected and followed by their subordinates.

(G) 3. "have done all right for themselves"; the ones who have managed to make enough money to live the way they want and who have good social and economic position.

(F) 4. say what they think, act as they feel and never do things just because others think them more proper.

Really enjoyable way to spend free time is—

(B) 5. being with family and friends. It isn't so much what you are doing, just as long as you can spend time with the people you like.

(C) 6. working on some hobby or skill, trying to improve your ability. Whether it's practicing golf strokes or working in a wood-working shop (or any similar activity) the important thing is to keep getting better and better at it.

(F) 7. "letting off some steam"; doing something where you can kind of let yourself go and not worry what others might think about it.

The pleasantest kinds of day-dreams are those concerning—

(A) 8. getting to be well known and recognized for something I have done which is really appreciated by others.

(G) 9. making money and getting into a good social position.

(K) 10. a long life of economic security, free from any worry about economic matters.

(D) 11. getting to be a stronger, more influential person, having more ability to direct others and have them follow me.

(I) 12. the heavy responsibilities somewhat taken off my shoulders; someone to do my worrying for me and to help me do the right things.

The things which must make being a child a very happy time are—

(K) 13. not having to worry about making a living. You know that your parents will support you and you don't have the economic worries of adults.

(B) 14. you get more real genuine love from your family than you ever get again in your life.

(I) 15. when you're a kid you can pal around with adults and older kids and can depend on them to help you out and teach you some things. When you're an adult you're more or less on your own and getting help from others is frowned upon.

(A) 16. getting to be recognized is easier in a group of children than in a group of adults. Sometimes you can really shine in the eyes of your friends by doing something that is really easy for you. It's much harder to do that when you're an adult.

When a man selects a hobby the important thing to consider is—

(C) 17. will it require enough skill so that he can really get some satisfaction from taxing his ability.

(E) 18. will his efforts produce anything worth-while in terms of happiness for others (e.g. fixing toys for orphan children before Christmas).

(F) 19. will it give him a chance to express himself.

(B) 20. will it give him a chance to make new acquaintances and meet new friends through interests in common things.

When I have a new job to do, something which I have never done before, I usually like to—

(L) 21. be left on my own to try it, whether or not I can do it correctly right away.

(J) 22. have to do some thinking and planning in an effort to figure out the right way to do it.

(I) 23. have somebody show me the right way to do it so I will be sure to do it right.

I admire and would like to be like the kind of man who

(H) 24. lives his life in accordance with a set of laws which always points out the right way of life for him.

(G) 25. always seems to make enough money to afford many things which I sometimes can't afford.

(J) 26. is always able to invent and create new things; who can meet new problems and solve them by his own skill.

(D) 27. seems to have the skill and knack for leading men and for getting others to follow his directions.

(E) 28. is always trying to help others; the man who gets his pleasure by giving happiness to others.

Assuming that I have my choice, I would want my epitaph to read somewhat like—

(E) 29. ". . . always interested in the welfare of his fellowman. His greatest satisfactions came with helping others."

(H) 30. ". . . lived in accordance with his ideals of righteousness. A man who had a credo and deviated not therefrom."

(L) 31. ". . . completely free. His life was not influenced by most of the pressures of life. He remained independent of . . ."

(A) 32. ". . . known and respected by many. His works were a constant source of admiration for him by the many who appreciated them."

(J) 33. ". . . blazing new frontiers by disregarding present methods and inventing new ones . . ."

If I were keeping a personal diary—one which nobody but me would ever see—the kind of entry which would give me satisfaction would be:

(K) 34. "At last I have a permanent job with a permanent income. Now, finally, I can relax and enjoy things without the constant worry about whether I'll be able to get a new job if the one I'm in finishes."

(C) 35. "Today I got transferred off that routine job. This one is tougher, but it sure feels good to have to try a little in order to do the work correctly."

(L) 36. "At last my supervisor isn't hanging around as much as he was at first. Boy, it sure feels swell not to have him around telling me how to do my work."

Be sure you have rated all 36 items in this part before going on to PART D.

PART D

DIRECTIONS. In this section you are asked to rate your satisfactions with *specific* aspects of your work. *DO NOT* consider your attitude toward your job as a whole (you have already done that in another part of the questionnaire). As you read each statement, think about how well you are satisfied with the specific item. Let your feelings be your guide in rating these items.

Use this scale:

(5) *completely* satisfied
(4) *very well* satisfied
(3) *well* satisfied
(2) *slightly* satisfied
(1) *not at all* satisfied

You can use *any* of the numbers as often as you like. *Rate every item.*

(A) 1. On my job when I do a piece of work I know that I'll get enough praise and recognition for it.

(B) 2. Where I work I get all the opportunity I want for making friends and enjoying the company of my fellow-workers.

(C) 3. When I've finished a day's work I can really be satisfied with the knowledge that I've used all my skills and abilities.

(D) 4. In the course of my work I have all the opportunity I might want to direct others.
(E) 5. My work results in benefits to many people.
(F) 6. My work offers me a real opportunity for self-expression.
(G) 7. The income I receive from my job enables me to live in a manner which I consider adequate.

(H) 8. I do not have to do anything on my job which is not in accordance with my ideas of right and wrong.

(I) 9. In my work I get all the help and supervision I need.
(J) 10. There is ample opportunity in my work to use my ingenuity and inventiveness.
(K) 11. I feel that my job is a secure one.
(L) 12. I have as much freedom as I want on my job.
(A) 13. In my work I always get the credit I deserve for any work I do.
(C) 14. I have to concentrate and put forth some effort to do my work, but it is not too hard for me.

(J) 15. I often have to think up some new ways of doing things and solving problems in the course of my work.

(B) 16. My job gives me plenty of opportunity to enjoy time with my family and friends.
(L) 17. On my job I am free from too much supervision.
(D) 18. I have as much responsibility as I want with respect to supervising the work of others.
(E) 19. My work is as worthwhile as most others I would want to be in with respect to helping other people.

(H) 20. In my job I am completely free of any worry about violating my religious or ethical values.
(F) 21. On my job I can always act just the way I picture myself—I don't have to act like somebody else.

(I) 22. I get all the help and advice that I need from my supervisors.
(K) 23. My job is quite permanent. It will be there as long as I might want it. If not, I at least know that I'll always have some sort of adequate income.
(G) 24. My present job enables me to have a good social standing.

PART E

DIRECTIONS. In this part you will be asked to *imagine yourself in various occupations.* With each occupation are several aspects of that occupation which may or may not contribute to the satisfaction of workers in that occupation. You are to rate each one on *how important* it would be for *your* satisfaction if *you* were in that occupation, according to the following scale:

(5) of *top importance* to my satisfaction
(4) *very important* to my satisfaction
(3) *quite important* for my satisfaction
(2) of *slight importance* to my satisfaction
(1) of *no importance* to my satisfaction

DO NOT consider your over-all liking or disliking of the occupation. Just take *each item separately* and rate it on whether or not it would be important to you if you were in the occupation. *Answer every item* even if you are not familiar with some of the occupations.

If I were a *policeman,* what I would like about the job is the fact that—

(K) 1. I would be a civil service employee and know that my job was secure and not subject to fluctuating business conditions.

(H) 2. the police enforce the laws, preventing those who might do wrong from harming other people.

(D) 3. the uniform of the police is a symbol of authority. When wearing that uniform I would know that my orders and directions would be followed.

(E) 4. the police department is an important agency in helping to alleviate suffering. The police are on hand at every disaster helping victims of accidents and fires and assisting the doctors.

If I were a *chemical engineer* what I'd like about it is the fact that I'd—

(J) 5. always have to face new and challenging problems for which I would always have to search for new solutions.

(C) 6. have fairly difficult work to do—difficult enough so that when I finished a job I could be pretty well self-satisfied with the job that I'd have done.

(A) 7. command respect and attention because of the high degree of skill required by the work. If I were to invent some new process or discover a new use for some substance my name would be associated with the discovery in all the professional journals and I would be respected throughout my profession.

If I were a *clergyman* (minister, priest or rabbi) what I would like about being in that profession would be that—

(H) 8. religious leaders are the ones who really show people how to lead the right kind of lives. They help bring religious teachings to the people and show them how to live according to their religions.

(D) 9. men in the clergy are leaders in their community. People look up to them and come to them for advice.

(E) 10. they give comfort and help to those who need it. They are a blessing to the needy.

If I were a *professional baseball player,* what I would like about that kind of work is the fact that—

(A) 11. there is a great deal of pleasure that comes with having a lot of fame. It must feel good to see your pictures in the newspapers and hear your name mentioned over the radio, and to have people stop you and ask for your autograph.

(G) 12. with the money I'd make in big-league playing, I'd really be able to live well. I could travel around staying in the nicest hotels, eating in the nicest restaurants without too much concern over how much it's all costing.

(B) 13. the team spirit that develops among members of the same team must really be strong. There is a good feeling that comes with being so close to your friends on the team and sharing your experiences with them.

(E) 14. a baseball player gives pleasure and happiness to thousands of people. It must make one feel very good to know that there might be thousands of kids in the grandstands who are having one of their happiest times watching the game.

If I were an *officer* in the *U.S. Army* the things that would contribute to my satisfactions in that job are the facts that—

(A) 15. military officers are people who are always looked up to and respected. Whether in a military or civilian group, the officer is recognized as being worth his salt.

(G) 16. as a commissioned officer with all the family allowances and other financial benefits one could live very comfortably and certainly very respectably. Most of the officers on a station are given nice homes to live in—and an officer of low rank might have a house as nice, or almost as nice, as one of the high-ranking officers.

(D) 17. there is a certain amount of satisfaction that comes with being in charge of a group of men—being able to direct them in their activities.

(H) 18. as an officer I could feel proud that my responsibility actually is the protection of our country.

If I were a *taxicab driver,* what I would like about the job would be the fact that—

(L) 19. I was my own boss, free to drive around wherever I wanted—choosing my own hours of work, and not having any boss around while I worked.

(F) 20. I could be myself. I could talk to the passengers and let them know what I think, and not have to behave in any special way.

(C) 21. you get to be such a skilled driver. A really experienced cab driver can handle a car perfectly. It's really something to be proud of.

(J) 22. there are so many times you have to use your brains to get out of a tough spot. There's no one around to tell you what to do. It's sometimes a real challenge to figure out your actions.

If I were a *laboratory technician* what would contribute to my satisfaction would be that—

(I) 23. I'd be working under scientists who knew much more than I and to whom I could really look to for advice, help, and leadership.

(C) 24. being a laboratory technician is really skilled and exacting work, and when you do your job properly you can really be proud of your work.

(J) 25. in a laboratory there are so many times when you really have to be inventive to be able to get your work done. There are lots of times you simply are forced to invent new ways of doing things. These problems would be exciting challenges.

If I were a *sailor* in the *U.S. Navy,* what would provide me my satisfaction on that job would be that—

(K) 26. no matter what happens you get your monthly pay check. It's always there for you no matter what happens. And then when you retire you know that you'll have your retirement pay.

(I) 27. you're never expected to do something on your own that you've never done before. If a new job ever comes up, the petty officer or one of the officers always shows you how to do it. There's always somebody around watching you so you never really have to take the responsibility if something goes wrong.

(B) 28. you make some very close friends in the Navy. Living close to your buddies the way you do there is a real chance for friendships and sharing of experiences.

If I were a *sculptor* what I would like about being in that field would be the fact that—

(J) 29. in art there is a great opportunity to create new and original ideas.

(B) 30. there is a common bond among artists and many friendships are established through a common interest and understanding.

(L) 31. a sculptor is free to live and work as he chooses. There is nobody to tell him how to work—he does so in whatever way pleases him.

(F) 32. In producing art works one can express one's feelings and emotions in a quite accepted way.

If I were a *motion picture producer* in Hollywood, what I would like about being in that kind of a job would be that—

(G) 33. such work would pay me enough to live a prosperous life. I could afford a beautiful home, an expensive car and all the other things that stand for success in one's business life.

(C) 34. whenever I completed a good picture I would get the satisfaction that comes with having done a difficult job well.

(D) 35. at times I would have as many as several hundred people under me and subject to my directions. Few people get such responsible positions.

If I were a *college professor,* what would make for my satisfaction in that job would be the fact that—

(L) 36. a teacher works on his own. He doesn't have to put up with constant supervision by his superiors.

(B) 37. the life-long friendships that teachers sometimes build up with their students are a constant source of satisfaction.

(K) 38. in most teaching jobs in colleges your appointment is quite permanent and you aren't likely to lose your job if you are doing satisfactory work.

(F) 39. teaching and lecturing allows a teacher to express his feelings and attitudes.

If I were an *apprentice shoe designer* what would be satisfying about that type of work would be that—

(F) 40. in that kind of work you have an ample opportunity for self-expression.

(I) 41. you could have all the help and advice you needed from the person for whom you worked.

(E) 42. you would be designing articles that would be useful and enjoyable for many people.

If I were a *cub reporter* assigned to work with a well-experienced reporter, what I would enjoy about such work would be the fact that—

(A) 43. a newspaper reporter has a certain status. Wherever a reporter goes, people recognize him by his profession and respect him for it.

(H) 44. in reporting work you are performing a real service to the concept of democracy by bringing enlightenment to the people.

(I) 45. I could work under somebody as skilled as that. I could learn from him and know how to do the right things by watching him so I could avoid stupid errors.

If I owned a large amount of stock in a very big and prosperous company and my only duties were to attend occasional shareholders meetings, what I would like about that way of life would be the fact that—

(L) 46. I would be completely independent and not have to take orders from anybody.

(K) 47. I would have a practically guaranteed income for the rest of my days and would never have to worry about it.

(G) 48. I could live in a manner which is so often portrayed in books and in the movies, yet which is so seldom enjoyed in real life—being able to afford the best of everything.

Be sure you have rated all 48 items in Part E.

REFERENCES

1. Ebel, R. L. Estimation of the reliability of ratings. *Psychometrika,* 1951, **16,** 407-424.
2. Friend, Jeannette G., & Haggard, E. A. Work adjustment in relation to family background. *Appl. Psychol. Monogr.,* 1948, No. 16.
3. Hendricks, I. Work and the pleasure principle. *Psychoanal. Quart.,* 1943, **12,** 311-329.
4. Hoppock, R., & Robinson, H. A. Job satisfaction researches of 1950. *Occupations,* 1951, **29,** 572-578.
5. Hoppock, R. *Job satisfaction.* New York: Harper, 1935.
6. Katz, D. *Human relations program of the Survey Research Center.* Ann Arbor: Univer. of Michigan, 1950.
7. Kornhauser, A. Psychological studies of employee attitudes. *J. consult. Psychol.,* 1944, **8,** 127-143.
8. Lewin, K. *Resolving social conflicts.* New York: Harper, 1948.
9. Murray, H. A. *Explorations in personality.* New York: Oxford Univer. Press, 1938.
10. Roe, Anne. Personality and vocation. *Trans. N. Y. Acad. Sci.,* 1947, **9,** 257-266.
11. Roethlisberger, F. J., & Dickson, W. J. *Management and the worker.* Cambridge: Harvard Univer. Press, 1939.
12. Schaffer, R. H. *Occupational attitudes survey.* New York: privately printed, 1950.
13. Schaffer, R. H. Job satisfaction as related to need satisfaction in work. Unpublished doctoral thesis, Teachers Coll., Columbia Univer., 1952.
14. Strong, E. K. *Vocational interests of men and women.* Palo Alto: Stanford Univer. Press, 1943.
15. Walker, Helen M., & Lev, J. *Statistical inference.* In press.

(Accepted for publication June 1, 1953)

Vol. 67, No. 15 Whole No. 365, 1953

Psychological Monographs: General and Applied

The Effects of Analeptic Drugs in Relieving Fatigue[1]

Robert H. Seashore
Northwestern University

and

A. C. Ivy
University of Illinois

In 1941-42 a series of studies was made of the effects of analeptic drugs on men engaged in prolonged activity. The work was prompted by the possible value of drugs in combating sleepiness, psychomotor inefficiency, and low motivation as might result from extended military activity. The results are now published since they are applicable to industrial and other phases of daily life wherein errors of human performance are to be minimized despite fatigue and prolonged wakefulness.

Physical activity in these studies consisted of, e.g., marches, guard duty, or truck and tank driving, together with a period of 24 hours of continuous wakefulness. The object of the studies was to determine the relative advantages and disadvantages of each of three analeptic drugs for use in large-scale military activity. None of the work tasks was simulated; they were the typical working tasks of the men in as nearly normal a context as possible. Since the primary concern was with the effects of the drugs on men in this actual working environment the tests and the investigators went into the "field" for the observations. The advantage of real working tasks and conditions placed limitations on the kinds and extent of observations. Tests were semiportable and only limited testing periods could be inserted into the work cycle of the men. With laboratory conditions and unlimited testing time the studies would have been expanded to the greater scientific satisfaction of the investigators but to the possible displeasure of the volunteer subjects. Since the completion of this work new technical information has been gained (19, 20, 21). Thus the testing procedures would now perhaps be modified in some details. The general approach, however, of integrating physiological, psychological, and medical investigation seems eminently successful. Furthermore, the experimental design and the results with their interpretation have applications beyond the military setting of the experiments.

---

[1] These studies were conducted with the assistance of James E. Birren, A. C. Van Dusen, and Stanley C. Harris, who administered the tests and assisted in the analysis and interpretation of the data. Henry C. Coopmans was instrument-maker and Claude E. Buxton aided in the design of the studies. It is impossible to thank individually the numerous volunteer subjects whose cooperation in the arduous routines made the studies possible. There were many who assisted in the arrangements for the studies whose valuable contributions must unfortunately be given an anonymous acknowledgment.

DESIGN OF PROJECTS

An attempt was made to set up fatigue studies under actual working conditions, including the usual rest and meal periods, that would approximate military operations covering a time period longer than the usual periods of wakefulness. In all but two of the experiments the subjects were soldiers, directed by field officers who were technically familiar with the special operating conditions. Six of the experiments used groups, averaging 16 men each, which were studied intensively four or more times; two mass experiments used groups of 250 men each.

An initial period of instruction and practice was given prior to each of the first experimental sessions. This was done in order to familiarize the subjects with the tests, to advance performance on the behavioral tests beyond the stages of rapid learning, and to relieve possible apprehension concerning the experiment.

Each project was called a "Vitamin Fatigue Project" in order to disguise the nature of the experimental substances. While it was indicated that the "vitamins" might aid the subject in remaining alert and relieving fatigue during prolonged duties, it was also indicated that there might be some minor deleterious side effects. The men were told that the projects were being made in order to learn which of the "vitamins" were likely to produce either good or bad effects and the extent to which they differed in these respects.

Drugs used and method of administration. The drugs used were caffeine sodium benzoate (7.5 gr.); Benzedrine Sulfate (10 mg.); desoxyephedrine hydrochloride (5 mg.); and a placebo of milk sugar (5 gr.). Each of the drugs was administered orally, in identical capsule form; the men were unable to distinguish between them.

The experimental design of the projects provided for division into squads and subgroups. Thus such factors as drug order, testing order, daily variation, etc. could be controlled or counterbalanced in such a way that the effectiveness of the drug could be statistically determined (see Table 1). Use of placebos permitted control of the psychological effects of any capsule administered.

Testing methods and instruments. All projects included a number of indices of effects of the drugs upon performance following fatiguing assignments. Some of the projects used all, others only part of the various tests. Tests were made as follows:

A. General Physiological Condition:

1. Diastolic and systolic blood pressure, average of two measurements each.
2. Pulse rate, checked until two identical 15-second counts were obtained.
3. Maximum time of breath-holding, average of 2 trials.
4. Visual flicker-fusion test. Average of three descending trials from continuous light to appearance of flicker, with the experimental light on for as short a time as possible during the test.
5. Itemized subjective reports on condition as to wakefulness, fatigue, aches and pains, bodily discomfort, etc.

B. Measures of Speed and Precision:

1. Visual discrimination reaction time (Seashore-Starmann instrument) (18). Subject places hands on telegraph keys, in front of each of which is a small ½-watt neon bulb. Instructions are to depress the key as rapidly as possible with the hand nearest to which the light appears. An average of 80 trials was used to measure speed.
2. Arm-hand steadiness in aiming (Seashore photoelectric target register) (19). The subject is seated, with elbow slightly away from his side and fingers extended horizontally holding a lever. By means of the lever he attempts to keep a small target directly in front of a light aperture, thus

controlling the light going to a photoelectric cell. In an oscillating circuit the photoelectric cell acts as a light-controlled resistance to an electric counter. The score is in terms of the sum of the distances multiplied by the times of each deviation from the true center. Measurement was made as the average score on 6 trials of 30 seconds each.

C. Measures of Tremor and Sway

1. Arm-hand tremor and sway (Seashore instrument) (20). The subject is seated as in the preceding aiming test, but holds a small ball attached to horizontal and vertical threads. The threads operate extremely lightweight ratchet mechanisms to give, on dials, the totals of movement up, down, right, and left. Six trials of 15 seconds each with eyes open were averaged.
2. Postural sway (Seashore instrument) (20).
    - *a*. Eyes open. Subject stands erect in a comfortable position, heels together, toes diverging at an angle of 45°. Threads from a light headgear operate lightweight ratchet mechanisms, recorded on dials, to give total sway front, back, right, and left. Two trials of 1 minute each were averaged.
    - *b*. Eyes closed. Same procedure as with eyes open.

D. Tridimensional Coordination

1. Miniature airplane test. The airplane rotates in three planes, each controlled by a separate lever or treadle. The left-hand lever controls horizontal rotation or steering (rudder); the right-hand lever controls banking movement of the wing; a right-foot treadle controls elevation (dive or climb) movements of the nose.

    During the test the airplane executes a continuous but irregular series of movements, induced by cam mechanisms. The subject attempts to counteract these movements in such a way as to keep the airplane in level flight and pointed straight ahead. Five trials of 2 minutes each were averaged.

E. Postexperimental Effects

The day after tests of the various drugs or placebo were made, each subject was given a simple report blank to fill out. The following questions were asked:

1. Did you have any trouble falling asleep?
2. Did you sleep well?
3. How long did you sleep?
4. How did you feel when you got up?

These questions were designed as a check on the possible influence of the drugs, both in preventing normal sleep and in producing effects noticeable after awakening on the following day.

THE PROJECTS

PROJECT I: FATIGUE FROM ALL-DAY MARCHES FOLLOWED BY ALL-NIGHT GUARD DUTY

Method. A voluntary group of 2 officers and 14 enlisted men was given initial tests of physical condition and muscular coordination (all of the experimental measures). They were then divided into two equal squads. Once a week for four weeks each squad participated in an all-day, 18–20-mile hike with 26-pound packs, followed by continuous guard duty from 6 P.M. until 3 or 4:30 A.M. Immediately thereafter the testing, which required 1½ hours, was started. Thus the subjects were on duty from the time of arising at 6 A.M. until 4:30 or 6 A.M. the following morning, approximately 24 hours of continuous wakefulness.

The tests were divided into four groups, each of which required 20 minutes. By prearranged schedule, each subject always took the series of tests in the same order. Principal details of the design of the experiment are given in Table 1.

A first capsule of either the placebo or one of the analeptic drugs (caffeine, Benzedrine, or desoxyephedrine) was given after supper at 6 P.M. and a second capsule of the same kind was given at midnight. By prearranged schedule, each subject received a different drug in each of his four weekly-interval experimental periods (see Table 1).

TABLE 1

DESIGN OF THE EXPERIMENT

Group	Subjects	Drug Order	Test Order	Weekly Assignment
1*A*	No. 1	1, 2, 3, 4	1, 2, 3, 4	7 A.M.–5 P.M.—hike
	“ 2	2, 3, 4, 1	2, 3, 4, 1	5 P.M.–6 P.M.—cleanup, eat
	“ 3	3, 4, 1, 2	3, 4, 1, 2	6 P.M.–3 A.M.—guard duty
	“ 4	4, 1, 2, 3	4, 1, 2, 3	3 A.M.–4:30 A.M.—testing
1*B*	No. 5	1, 2, 3, 4	3, 4, 1, 2	7 A.M.–5 P.M.—hike
	“ 6	2, 3, 4, 1	4, 1, 2, 3	5 P.M.–6 P.M.—cleanup, eat
	“ 7	3, 4, 1, 2	1, 2, 3, 4	6 P.M.–4:30 A.M.—guard duty
	“ 8	4, 1, 3, 2	2, 3, 4, 1	4:30 A.M.—6 A.M.—testing
2*A*	No. 9	1, 2, 3, 4	2, 3, 4, 1	7 A.M.–5 P.M.—hike
	“ 10	2, 3, 4, 1	3, 4, 1, 2	5 P.M.–6 P.M.—cleanup, eat
	“ 11	3; 4, 1, 2	4, 1, 2, 3	6 P.M.–3 A.M.—guard duty
	“ 12	4, 1, 2, 3	1, 2, 3, 4	3 A.M.–4:30 A.M.—testing
2*B*	No. 13	1, 2, 3, 4	4, 1, 2, 3	7 A.M.–5 P.M.—hike
	“ 14	2, 3, 4, 1	1, 2, 3, 4	5 P.M.–6 P.M.—cleanup, eat
	“ 15	3, 4, 1, 2	2, 3, 4, 1	6 P.M.–4:30 A.M.—guard duty
	“ 16	4, 1, 3, 2	3, 4, 1, 2	4:30 A.M.–6 A.M.—testing

Because the drugs ordinarily have a latent time of approximately 30–45 minutes and an effective period of about 7–8 hours, it was expected that the men who were tested at 3 A.M. would show maximal effects of the second dose, while those who were tested at 4:30 A.M. would show the later stages of effects of the drugs.

Subjective results. As will be noted in the summary of subjective reports given for each drug (Table 2), the subject's own reports were better for all three of

TABLE 2

COMPARISON OF SUBJECTIVE REPORTS FOR EACH DRUG

Item	Measure	Placebo	Caffeine	Benzedrine	Desoxy-ephedrine
Hours sleep night before experiment	Average.....	7.0	7.0	6.8	7.1
	Median.....	7	7	7	7
Estimated number of hours subject could have continued guard duty	Average.....	1.9	4.2	4.2	2.8
	Median.....	1	2	4	2
“Capsule helped” (no. subjects)		6	12	14	12
“Capsule no effect” (no. subjects)		7	2	0	3
“Capsule hindered” (no. subjects)		2	0	0	0
Legs: Tired		2	2	6	6
Sore-aching		3	4	3	8
Pain		4	2	0	1
	Total.......	9	8	9	15
Feet: Tired		1	1	1	2
Sore-aching		3	1	1	6
Pain		8	6	3	3
	Total.......	12	8	5	11
General condition of subject: Good		5	11	14	13
Poor		10	3	0	2

TABLE 3

COMPARISON OF NEGATIVE SUBJECTIVE SYMPTOMS FOR EACH DRUG

Symptom	Placebo	Caffeine	Benzedrine	Desoxy-ephedrine
Excessive sleepiness	12	5	3	9
Visual or auditory impairment	9	6	5	9
Vertigo	4	2	2	3
Easily fatigued	9	2	1	6
Feeling of heat and sweating	2	0	4	6
Tremors	7	9	6	11
Impaired coordinations	7	4	2	8
Palpitations or cardiac distress	3	2	2	1
Stuttering or blocking of speech	1	1	1	4
Difficulty in concentrating	5	1	1	2
Slowness in reasoning	5	1	1	4
Greater effort to carry out tasks	12	1	0	8
Mentally lazy	9	1	0	2
Depressed and grouchy	7	1	0	4
Nervous, inward tension	6	2	0	4
Fidgety or restless	7	3	2	4
Indifferent and exhausted	8	2	0	5
Total	113	43	30	90

the stimulant drugs than for the placebo; Benzedrine was given a clear preponderance of favorable reports over each of the other stimulant drugs.

In this connection, it was of interest to note that the subjects could tell rather accurately when they had been given a control dose of "dud," as they described it. The experimenters, too, were able to recognize fairly accurately when the subjects had received the placebo, merely by sleepy behavior exhibited during the test period. They were also often able, by the more buoyant attitude and behavior exhibited, to distinguish subjects receiving Benzedrine from those receiving the other two stimulants.

Negative subjective symptoms, summarized in Table 3, showed that the drugs differed significantly among themselves, with Benzedrine ranking first in freedom from negative symptoms, followed closely by caffeine and markedly separated from desoxyephedrine, which in turn differed only moderately from the placebo.

Positive subjective symptoms, summarized in Table 4, showed Benzedrine to be first in desirable effects, followed by caffeine, desoxyephedrine, and the placebo.

Objective results. Results from the administration of objective physiological and behavioral measurements are summarized in Table 5. The probabilities, p, in Table 5 were derived from the t test of significance for matched samples (4). The p values indicate the percentage of future cases in which such a difference between means would likely occur by

TABLE 4

COMPARISON OF POSITIVE SUBJECTIVE SYMPTOMS FOR EACH DRUG

Symptom	Placebo	Caffeine	Benzedrine	Desoxyephedrine
No sleepiness	3	9	11	6
Talkative and excited	1	7	10	5
Feeling of exhilaration	1	5	10	4
Total	5	21	31	15

TABLE 5

PERCENTAGE OF ABSOLUTE DIFFERENCES BETWEEN MEAN RAW SCORE UNDER PLACEBO AND EACH DRUG CONDITION, WITH p INDICES FOR STATISTICAL SIGNIFICANCE OF DIFFERENCE OF RAW AND z SCORES*

Measure or Test	Caffeine vs. Placebo			Benzedrine vs. Placebo			Desoxyephedrine vs. Placebo		
	% abso. dif. in av. raw scores	p indices		% abso. dif. in av. raw scores	p indices		% abso. dif. in av. raw scores	p indices	
		Raw score	z-score		Raw score	z-score		Raw score	z-score
Pulse	2	50–60	20–30	17	*1*	*1*	9	5–10	*1–2*
Systolic blood pressure	3	5–10	30	5	*2–5*	*2–5*	3	10–20	20–30
Diastolic blood pressure	7	*2–5*	*1–2*	8	*1*	*1–2*	4	10–20	10–20
Pulse pressure	4	10–20	30–40	0	90–100	80–90	1	80–90	60
Breath-holding	8	20–30	10–20	9	20–30	10–20	10	*2–5*	*1*
Flicker-fusion	5	10–20	30–40	7	*1*	*2–5*	3	10–20	50–60
Discriminatory reaction time	5	10–20	10–20	6	*1*	5–10	6	*1*	*2–5*
Steadiness in aiming	8	30–40	5–10	6	50–60	10–20	7	30–40	10–20
Tridimensional coordination (all)	15	10–20	*2–5*	31	*1*	*1*	22	*2–5*	*1*
Tridimensional coordination (4+5)/2	22	5–10	*2–5*	42	*1*	*1*	26	*1*	*1*
Hand-arm sway	5	60–70	80–90	20	*2–5*	5–10	18	*2–5*	5–10
Body sway, eyes closed	0	90–100	80–90	5	60–70	40–50	11	20–30	50–60
Body sway, eyes open	3	60–70	40–50	12	5–10	*1–2*	11	*2–5*	*1–2*
Medians	5	10–20	20–30	8	*2–5*	*2–5*	9	*2–5*	*1–2*

* Statistically significant values, i.e., $p < .05$ are in italics.

chance alone.

Two such indices are given for each drug, as compared with the placebo: (*a*) comparison of average raw score of all subjects at the time of taking the particular drug, without regard to the week of the experiment, and (*b*) in terms of z scores, or the standard deviations above or below the mean, of all persons taking the test during the week for which each comparison was made. This second index corrects for the fact that some of the tests showed significant continued effects of practice. Where such practice effects were fairly large, the second index usually shows greater significance, but where little improvement resulted from practice, the statistical treatment sometimes showed slight reversals of significance (probably attributable to the small number of subjects studied).

In examining Table 5, it will be seen that all the behavioral tests showed indices favoring use of the three stimulants, ranging from moderate to high degrees of statistical significance. Breath-holding, a measure of motivation and, to some degree, of specialized strength, showed a moderate improvement for all three stimulants.

Visual flicker-fusion, often considered to be primarily an indicator of neural fatigue or recovery, showed smaller but consistent improvement for Benzedrine, lesser improvements for the other two stimulants. Speed of visual discriminative reaction time was slightly but reliably improved by all three stimulants. Hand-arm sway was markedly improved for Benzedrine and desoxyephedrine, slightly improved for caffeine. To a lesser extent the same was true for postural sway with eyes open.

Postexperimental results. A summary of results is shown in Table 6. The average number of hours of sleep

TABLE 6

SUBJECTIVE REPORTS ON SLEEP AND CONDITION FOLLOWING EXPERIMENT*

Subjective Report	Placebo	Caffeine	Benzedrine	Desoxy-ephedrine
Average no. of hours sleep	4.7	4.5	4.0	4.8
Median no. of hours sleep	4.5	4	4	5
Trouble falling asleep (no. of subj.)	0	4	3	1
Good restful sleep (no. of subj.)	13	10	11	12
Good condition after sleep (no. of subj.)	10	8	10	7
Poor condition after sleep (no. of subj.)	4	5	4	7

* Benzedrine was the only drug for which there was a report from every subject.

(slightly over four) following the experiment was surprisingly short. This was attributed to the fact that the subjects were accustomed to late night hours and early rising, and to the noisy daytime conditions in the barracks.

No significant difference was found among the stimulants or the placebo with respect to average length of sleep following the experiment. The stimulants did not differ significantly from one another, or from the placebo, in quality of sleep or as to general condition after sleep.

Summary of results. Of the three stimulants, in general the best (subjective, objective, and postexperimental) results were obtained with Benzedrine. Caffeine produced desirable subjective reports with but little improvement in motor performance. Desoxyephedrine produced little change in subjective reports but produced improvements in motor effects that were as good as, or slightly better than, those with Benzedrine.

PROJECT II: BATTALION MARCH EXPERIMENT

Method. Four batteries of soldiers of about 70 men each participated in two regular training marches of 10–12 miles each on successive afternoons. The marches were largely on paved roads, over slightly rolling country, with a little cross-country work and in warm, humid weather. Equipment carried weighed approximately 26 pounds.

The three stimulants and a placebo, as used in Project I, were administered to two different batteries of men as follows:

Battery	First March No.	Second March No.
E	1—placebo	3—Benzedrine
F	2—caffeine	4—desoxyephedrine
G	3—Benzedrine	1—placebo
H	4—desoxyephedrine	2—caffeine

At the end of each march each man filled out individual cards which reported subjective effects. The more extended battery of sensorimotor tests was not used in this study. The number of men reporting for each drug was as follows: placebo, 131; caffeine, 148; Benzedrine, 137; desoxyephedrine, 147.

Subjective results. Since each drug was administered to approximately equal numbers of subjects on both marches, variation between marches was counterbalanced and should not affect interpretation.

Ranking the drugs according to favorable effect on the subjects, as indicated by responses on the questionnaire card, yielded the following results:

	Drug Order
Least difficulty with march	3, 4, 2, 1.
Least foot trouble	3, 4, 2, 1.
Least leg trouble	4, 3, 1, 2.
Least general fatigue	3, 1, 2, 4.
Greater "help" from capsule	4, 3, 2, 1.

Assigning equal weights to each of the

above items, the over-all order of merit among the drugs would be 3 (Benzedrine), 4 (desoxyephedrine), 2 (caffeine), and 1 (placebo). This finding agrees with those of Project I, which showed Benzedrine to produce the best subjective effects, but reverses the order between caffeine and desoxyephedrine.

Summary of results. Both Benzedrine and desoxyephedrine appeared to produce effects which were large enough to be recognized by untrained soldiers during a march. None of the stimulants showed any greater percentage of ill effects than did the placebo. The fact that at least 15 per cent of the subjects taking the placebo reported beneficial effects indicates the significance of suggestion in this as in many other medical treatments.

PROJECT III: COMMERCIAL TRUCK DRIVERS EXPERIMENT

Method. Drivers of large semi-trailer trucks, following their usual runs, were used for the experiment. Most of their work, including loading boxes of groceries, driving, and unloading, was done at night and alone. The total number of working hours ranged from 6 to 15 hours, with brief opportunities for time out for meals. Each driver was tested following one of his normal cross-country trips.

Before each trip an envelope containing one of the three stimulants or the placebo, as used in previous experiments, was issued to the driver with written instructions concerning the time (5–6 hours prior to end of run) to take the capsule. In every case the driver signed a slip giving the time he took the capsule.

Results. Relative shortness of most of the routes and inability to maintain fixed schedules because of mechanical and other difficulties prevented fixed intervals between drug ingestion and testing and precluded a detailed analysis of findings. This experiment did establish that a considerably longer (6–15 hours) and more fatiguing schedule would be required before effects of the drugs could be properly evaluated on truck drivers.

In general, while subjective effects of the stimulants were desirable and no important undesirable side effects were produced, there would probably be relatively slight benefits from routine use of the drugs by drivers on ordinary short schedules.

PROJECT IV: ARMY TRUCK DRIVERS EXPERIMENT

Method. Two groups, each consisting of an officer and eight drivers, worked on alternate days driving 2½-ton heavy-duty army trucks from 7 A.M. to 12 noon, from 1 to 5 P.M., from 6 P.M. to midnight, and, after a light lunch, from 12:15 to either 3 or 4:30 A.M. A first capsule of either caffeine, Benzedrine, desoxyephedrine, or the placebo was given after supper at 6 P.M. and a second capsule of the same kind was administered directly after the midnight lunch. Drives were made under a variety of conditions, including three hours of "blackout" on very poor country roads.

Immediately following completion of the drives, essentially the same subjective, physiological, and sensorimotor behavioral tests as used in Project I were made.

Subjective results. There were relatively few negative subjective symptoms associated with any of the stimulants and no significant differences between them. Among the positive subjective findings, Benzedrine was reported as eliminating sleepiness in over 50 per cent; caffeine had a slight effect in a few subjects. All three of the stimulants exceeded the placebo in respect to the estimate of additional time subjects thought they could continue the same work.

Objective results. Physiological measurements were statistically significant in only two instances: Benzedrine produced an average increase of 18 per cent in the pulse rate and desoxyephedrine produced a 12 per cent increase. Since the normal pulse rate for these subjects was lower than that of the general population, these increases are not viewed as serious. Also, none of the other studies with Benzedrine produced such a pulse rate increase.

Flicker-fusion produced small but statistically significant improvement and the photoelectric aiming test produced a slightly greater degree of improvement with all three stimulants. Reaction time showed small and moderately significant differences in favor of Benzedrine and desoxyephedrine. Hand-arm sway and postural sway with eyes open showed only moderately significant differences in favor of all three stimulants; postural sway with eyes closed showed larger and more significant differences for all three.

Postexperimental results. Only two subjects, one of whom had desoxyephedrine, the other caffeine, had any difficulty in falling asleep. The only subject reporting feeling bad on awakening had received a placebo.

Summary of results. This experiment produced less observable or measurable degrees of fatigue and therefore fewer possible advantages attributable to the stimulants. What advantages did appear, however, were in line with the findings on the subjects in Project I.

PROJECT V: TANK DRIVERS EXPERIMENT

Method. Four tank crews, each consisting of four men, worked for five-hour periods on five successive days. Each man served as driver for 2½ hours, in rotation, thus driving only every other day. Since previous experiments seemed to indicate that in general Benzedrine was the drug of choice among the three stimulants, this experiment was limited to a comparison of Benzedrine with the placebo. Each member of a crew received the same drug, the capsule being administered just prior to the five-hour drive. Crews were tested immediately upon completion of the runs.

Subjective results. The only negative subjective symptom of even approximately statistical significance was the greater number of mild headaches and complaints about excessive dust under placebo conditions. Among the positive subjective effects Benzedrine showed only moderate advantages in avoiding sleepiness during the experiment and in reducing complaints about muscular conditions.

Objective results. All the sensorimotor tests showed relatively slight absolute improvements under Benzedrine, with moderately significant improvements in reaction time, photoelectric aiming, and postural sway with eyes closed.

Summary of results. Considering the short work periods employed, the most significant findings in this experiment were that administration of the drug under the extreme heat, vibration, and other aspects of tank operation did not produce undesirable side-effects and that use of Benzedrine did produce moderately desirable effects.

PROJECT VI: EFFECTS OF HOT-MOIST AND HOT-DRY TEMPERATURES UPON PHYSIOLOGICAL AND PSYCHOLOGICAL FUNCTIONS

At the time these experiments were made there was a current belief that Benzedrine might inhibit perspiration. Since it was anticipated that use of the drug might be used under desert and tropical conditions, it was important to find out if normal body-cooling mechanisms would be inhibited by use of the drug under hot-dry or hot-moist conditions.

Method. Eight students (six from medical schools) in groups of four each served as paid subjects. On six successive days sensorimotor performances were measured prior to and following exposure to high temperature conditions. No physical work was done. Each group was given the preliminary series of tests late in the morning, followed by a light lunch and administration of either a Benzedrine or a placebo capsule. They were then placed in one of the controlled temperature rooms for a period of six hours.

In the hot-moist room, the average temperature was 86° F. dry bulb and 85° F. wet bulb, with a relative humidity so high that the air was usually foggy. In the hot-dry room, the average temperature was 117° F. dry bulb, 85° F. wet bulb, with a relative humidity of approximately 17 per cent. During the six hours spent in the controlled temperature rooms, the previously trained medical students in the group took rectal temperatures and pulse and blood pressure readings on all subjects at intervals of approximately 40 minutes.

Summary of results. Six hours of exposure to hot-moist and to hot-dry conditions did not produce evidence of undesirable effects from either Benzedrine or the placebo. Operation of the body-cooling mechanism was adequate under both, as shown by visible evidence of perspiration under hot-moist conditions, by a very heavy water and salt intake under hot-dry conditions, and by absence of any unusual changes in rectal temperature, pulse rate, or blood pressure. The slight differences shown in sensorimotor measurements were, in nearly all cases, in favor of the use of Benzedrine.

PROJECT VII: DESERT INFANTRY MARCH EXPERIMENT

Method. Since night marches are the typical operation under desert warfare conditions, marches were started at either 7 or 8 P.M. Two officers and 16 enlisted men participated in each of four experiments. The first two experiments consisted of 7 P.M. to 6 A.M. marches of 17–22 miles, with 8-pound packs, followed by guard duty until either 10 or 11 A.M. When these experiments did not prove to be sufficiently fatiguing, the guard duty was changed to additional marching (making a total of about 26–28 miles) for the second two experiments.

For the first two experiments only one capsule of either Benzedrine or the placebo was administered directly after the midnight lunch at 12:30 A.M. On the third and fourth marches, a second capsule of the same kind was administered at 6 A.M., just prior to breakfast. Subjects reported for the test program at either 10 or 11 A.M.

Subjective and postexperimental results. No statistically significant differences between Benzedrine and the placebo were shown for either negative or positive subjective symptoms or for records of sleep following the experimental periods. Of the few complaints of tired muscles, much smaller percentages were reported following Benzedrine administration.

Objective results. Of the physiological measures of circulatory conditions, Benzedrine gave the greatest score in each case, although absolute changes were quite small. The best score on each of the behavioral tests was in favor of Benzedrine, although only in body sway with eyes opened and closed and in flicker-fusion were improvements of statistical significance.

Summary of results. In general, there

appeared a tendency toward improvement in performance under the Benzedrine condition, even though absolute changes in many instances did not reach statistical significance for either the subjective reports or objective tests. Positive influence of the Benzedrine was more apparent under the third and fourth experiments which were more fatiguing than were the first two.

PROJECT VIII: DESERT TANK EXPERIMENT

Method. Four groups of four men each served on alternate days for one pretesting period and four experimental sessions. The first two experiments used six hours of driving time, the third and fourth experiments used seven hours of driving time. Tests were made after completion of the drives. One capsule of either Benzedrine or the placebo was administered 4½–5 hours before testing.

Subjective and postexperimental results. There were no significant differences between Benzedrine and the placebo for either negative or positive subjective symptoms. Only two negative reports on sleep following the experiment were obtained, each of which was under the placebo condition.

Objective results. Absolute changes in circulatory condition were of no practical significance. Although better scores were made under the Benzedrine condition in all but one (discriminative reaction time) of the objective tests, absolute differences were so small as to be of no statistical significance.

DISCUSSION

Since too little is known about the fundamental nature of fatigue and recovery, it is impossible to interpret in detail the present findings. Results of behavioral studies such as the present will, of course, eventually need to be traced back to their detailed physiological and biochemical bases. Since these studies were conducted, new evidence is available about the effects of stimulants on the human subjects (1, 2, 3, 5, 6, 8, 9, 10, 11, 24, 25). They point, as did the present work, to the subtle nature of fatigue and to the necessity for having the subjects perform for periods much longer than conventionally regarded as fatiguing before drug effects may be demonstrated reliably.

The physiological effects of fatigue may be such that larger absolute values or larger relative differences in stimuli become necessary in order to produce the same effectiveness of response. The stimulant drugs used in these experiments were observed to overcome boredom and to produce a more alert attitude in the subjects. This may have resulted from an influence on the background stimulation. Use of stimulants may improve the performance of a task by facilitating the direction of attention to stimuli related to the main task. Thus the subject may discover new leads to effective action such as short cuts, or detect incipient errors which might lead to danger if uncorrected.

In evaluating the preceding experimental results, certain implications of the method employed should be recognized. Since we usually compared end results following use of a stimulant with those following use of a placebo, the differences measured represent only a part of the total fatigue, the remainder being effective under both stimulant and placebo conditions. Thus our measurements of differences in fatigue are conservative indices. Also, since we studied, primarily, differences between, and results of, stimulant and placebo conditions, we did not measure psychological benefits of drug administration since they were also present in the placebo.

Deleterious side effects or aftereffects of the stimulants were notably absent. It is probable that this was due in part to the moderate dosages used. It is also possible that heavy work, such as was performed, resulted in rapid, complete metabolism of the drugs. Considering evidence from the experiments with the battalion of men on training marches and the more intensive studies made with groups of about 16 men each, it seems reasonable to anticipate no greater ill effects from these drugs than would be expected by chance from other mild medical treatments—so few, in fact, as to be negligible.

With the possible exception of the first experiment, only relatively slight degrees of fatigue were observed or measured during the entire project. It was gratifying to learn that the men were in sufficiently good condition to carry out the experiments—which civilians and officers alike had estimated as near the limit of capacity—with such relatively slight decrement in well-being and efficiency. However, since, in general, the most significant benefits from use of the stimulants appeared in those experiments where observable fatigue was greatest, a more clear-cut demonstration of statistical significance would probably have resulted from use of subjects exhibiting greater degrees of fatigue.

Previous laboratory experiments with Benzedrine showed relatively little effect on sensorimotor coordinations when only slight fatigue was involved (4). This series of experiments, also, demonstrated limitation of benefit from stimulants under conditions of only slight fatigue. Therefore it would seem to follow that, except under certain conditions, stimulants should be administered only when fatigue is greater than that resulting from normal work. Exceptions would include such conditions as standing guard or driving at night, where alertness is important. In this connection, it should be recognized that, in the vast majority of comparisons made between the stimulants and the placebo in this study, differences were consistently in favor of the stimulants.

Medical Comment and an Opinion

If tired troops, who must serve regardless of fatigue, cannot get coffee, a stimulant is indicated. Benzedrine will keep them awake and alert as does coffee but, unlike the occasional effect of coffee, will not make their muscular action unsteady and jerky. Since it is not metabolized as rapidly as caffeine, the effect of Benzedrine should last longer.

We have seen no deleterious effects or objectionable side-reactions in 75 male medical students given 10 mg. of Benzedrine orally. Most reported that, 6–8 hours after taking the dose, they fell asleep readily after going to bed. In the physically fatigued soldiers used in these experiments, no significant disturbance of sleep was observed and no deleterious effects from Benzedrine were noted.

Benzedrine is not a substitute for sleep and rest; it should be used for emergencies only and its distribution should be kept under supervision.

In tired men or in men doing physical work, 10 mg. Benzedrine every six hours for two or three doses is not excessive. In light work, 5 mg. as a first dose, to be repeated in one or two hours, is perhaps preferable, although we have not given the 10 mg. in such divided doses. The 5-mg. dosage would be less likely to harm the very occasional subject with idiosyncrasy.

SUMMARY

A series of experiments was made by Northwestern University investigators for the Committee on Medical Research of the Office of Scientific Research and Development. Eight projects were used to study the effects of currently used stimulant drugs in promoting wakefulness and improved motivation, as well as their effects in overcoming fatigue during prolonged activity.

The drugs, administered orally in identical capsule form, were: caffeine sodium benzoate (7.5 gr.) Benzedrine Sulfate (10 mg.); desoxyephedrine hydrochloride (5 mg.); and a placebo of milk sugar (5 gr.). The drugs were always called "vitamins" and the projects "Vitamin Fatigue Projects" so that the nature of the experimental material was unknown to the subjects.

Records and tests included the following:

1. General physiological condition: systolic and diastolic blood pressure; pulse rate; breath-holding; visual flicker test; itemized subjective report.
2. Measures of speed and precision: visual discriminative reaction time; arm-hand steadiness.
3. Measures of tremor and sway; postural sway with eyes open and closed.
4. Tridimensional coordination.
5. Postexperimental subjective reports.

In the eight experimental projects, six groups, averaging about 16 subjects each, were used, and in one project four groups of about 70 subjects each were used. In the first four projects effects of all three of the stimulants were compared with the placebo. Because Benzedrine appeared to give superior results, only this drug was used for comparison with the placebo in the last four projects.

Project I produced degrees of fatigue that were more observable and measurable than those of the other seven projects. Project I is reported in detail, the other projects are merely summarized. A summary of significant results of subjective reports and objective tests is given in the appendices.

In general it may be stated that comparison of results, after using each of the three stimulants, with those after using the placebo, showed the following:

Caffeine: desirable subjective effects; negligible improvements in sensorimotor coordinations.

Desoxyephedrine: insignificant subjective improvements; significant improvements on sensorimotor measurements.

Benzedrine: subjective effects as good as those of caffeine; sensorimotor improvements nearly as great as those with desoxyephedrine.

None of the stimulants showed more than relatively unimportant side-effects.

APPENDIX 1-A

NUMBER OF SUBJECTIVE REPORTS UNDER PLACEBO AND BENZEDRINE

Subjective Report	Project I March & Guard		Project IV Truck Drivers		Project V Tank Crews		Project VII Marches		Project VIII Tank Crews	
	Plac.	Benz.	Plac.	Benz.	Plac.	Benz.	Plac.	Benz.	Plac.	Benz.
Positive symptoms....	2	31	1	4	4	8	6	9	5	3
Negative symptoms...	113	30	27	23	124	102	36	41	20	13
Tired muscles........	12	5			11	3	12	7	0	2
Headache:										
Yes...............			0	2	6	1	3	3	1	1
No................			16	14	33	38	32	33	29	31
Capsule:										
Helped............	6	14	11	13	10	12	21	27	18	25
Hindered..........	7	0	2	2	2	0	0	0	0	0
No effect..........	2	0	3	3	26	27	12	8	12	6

APPENDIX 1-B

NUMBER OF SUBJECTIVE REPORTS UNDER PLACEBO AND UNDER CAFFEINE AND DESOXYEPHEDRINE

Subjective Report	Project I March and Guard			Project IV Truck Drivers		
	Placebo	Caffeine	Desoxy-ephedrine	Placebo	Caffeine	Desoxy-ephedrine
Positive symptoms..........	2	12	9	1	5	5
Negative symptoms..........	113	43	90	27	25	25
Tired muscles..............	12	8	11	5	5	5
Capsule:						
Helped.................	6	12	12	11	13	15
Hindered................	7	2	3	2	0	0
No effect...............	2	0	0	3	2	1

APPENDIX 2-A

PERCENTAGE OF ABSOLUTE DIFFERENCE IN OBJECTIVE TEST SCORES UNDER PLACEBO AND BENZEDRINE CONDITIONS

Measure or Test	Project I March and Guard	Project IV Truck Drivers	Project V Tank Crews	Project VI Dry-Hot Temp.	Project VI Wet-Hot Temp.	Project VII Marches	Project VIII Tank Crews
Pulse rate................	17*	18*					8*
Systolic blood pressure.....	5*	4	5	12	2	6	
Diastolic blood pressure....	8*	5	2	2	2	3*	2
Pulse pressure............	0	0	5*	3*	4	4*	2
Breath-holding............	9	14	4	3	0	3	4
Flicker-fusion.............	7*	3*	2	6	9		
Discriminative reaction time	6*	5	1	15	1	2*	1
Photoelectric aiming.......	6	8	3	3	4	1	4
Tridimensional pursuit†	31*	2	5*	4	10*	2	2
Hand-arm sway...........	20*	20*	15				
Postural sway:							
1. Eyes closed..........	5	26*	14*	12*	9*	7	1
2. Eyes open...........	12	14	1	8*	8	8*	3
Serial discrimeter..........						1	6
Median..............	7.5*	8	4.5	7	8.5	3.5	4

* Statistically signifiant (p values .01–.05).

† z-score differences given here because of learning function.

APPENDIX 2-B

Percentage of Absolute Difference in Objective Test Scores under Placebo and Caffeine or Desoxyephedrine

Measure or Test	Project I March and Guard		Project IV Truck Drivers	
	Caffeine	Desoxy-ephedrine	Caffeine	Desoxy-ephedrine
Pulse rate	2	9	2	12*
Systolic blood pressure	3	3	3	5
Diastolic blood pressure	7*	4	3	2
Pulse pressure	4	1	3	5
Breath-holding	8	10*	10	6
Flicker-fusion	5	3	3*	4*
Discriminative reaction time	5	6*	1	3*
Photoelectric aiming	8	7	5	10*
Tridimensional pursuit†	15*	22*	23*	10
Hand-arm sway	5	18*	8	11
Body sway:				
1. Eyes open	0	11	19*	23*
2. Eyes closed	3	11*	13	27*
Median	5	9*	4	8

* Statistically significant (p values .01–.05).

† z-score difference given here because of learning function.

REFERENCES

1. Alles, G., & Feigen, G. Influence of Benzedrine on work decrement and patellar reflex. *Amer. J. Physiol.*, 1942, **136**, 392-400.
2. Alwell, N. Frequency and duration of the subjective and secondary effects of Benzedrine and pervitin on intensely fatigued persons. *Acta Med. Scand.*, 1943, **114**, 6-32.
3. Cuthbertson, D. P., & Knox, J. A. C. The effects of analeptics on the fatigued subject. *J. Physiol.*, 1947, **106**, 42-58.
4. Edwards, A. L. *Experimental design in psychological research.* New York: Rinehart, 1950.
5. Foltz, E., Schiffrin, M., & Ivy, A. The influence of amphetamine (benzedrine) sulfate and caffeine on the performance of rapidly exhausting work by untrained subjects. *J. Lab. clin. Med.*, 1943, **28**, 601-603.
6. Foltz, E., Ivy, A., & Barborka, C. The influence of amphetamine (benzedrine) sulfate, d-desoxyephedrine hydrochloride (pervitin) and caffeine upon work output and recovery when rapidly exhausting work is done by trained subjects. *J. Lab. clin. Med.*, 1943, **28**, 603-606.
7. Jones, H. E., & Seashore, R. H. The development of fine motor and mechanical abilities. *Yearb. nat. Soc. Stud. Educ.*, 1944, **43**, 123-145.
8. Kleemeier, Lyla Bechtel, & Kleemeier, R. W. Effects of benzedrine sulfate (amphetamine) on psychomotor performance. *Amer. J. Psychol.*, 1947, **60**, 89-100.
9. Knoefel, P. K. The influence of phenisopropyl amine and phenisopropyl methyl amine on work output. *Proc. Cent. Soc. clin. Res.*, 1942, **15**, 86-87.
10. Newman, H. W. The effect of amphetamine sulfate on performance of normal and fatigued subjects. *J. Pharm. exp. Therap.*, 1947, **89**, 106-108.
11. Pellmont. Comparison of the effects of coramine, caffeine and pervitin on mental and physical performance of fatigued and non-fatigued human subjects. *Arch. exp. Pathol. Pharmacol.*, 1942, **199**, 274-291.
12. Seashore, R. H. Stanford motor skills unit. In W. R. Miles and D. Starch (Eds.), University of Iowa studies in psychology. No. XII. *Psychol. Monogr.*, 1928, **39**, No. 2 (Whole No. 178), 51-66.
13. Seashore, R. H. Individual differences in motor skills. *J. gen. Psychol.*, 1930, **3**, 38-65.
14. Seashore, R. H., & Adams, R. D. The measurement of steadiness: a new apparatus and results on marksmanship. *Science*, 1933, **78**, 285-287.
15. Seashore, R. H. An experimental and theo-

retical analysis of fine motor skills. *Amer. J. Psychol.*, 1940, **53**, 86-98.

16. Seashore, R. H., Buxton, C. E., & McCollom, J. N. Multiple factorial analysis of fine motor skills. *Amer. J. Psychol.*, 1940, **53**, 251-259.
17. Seashore, R. H., & Seashore, S. H. Individual differences in simple auditory reaction times of hands, feet, and jaws. *J. exp. Psychol.*, 1941, **29**, 342-345.
18. Seashore, R. H., Starmann, R., Kendall, W. E., & Helmick, J. S. Group factors in simple and discriminative reaction times. *J. exp. Psychol.*, 1941, **29**, 346-349.
19. Seashore, R. H., Dudek, F. J., & Holtzman, W. A factorial analysis of arm-hand precision tests. *J. appl. Psychol.*, 1949, **33**, 579-584.
20. Seashore, R. H., & Dudek, F. J. A factorial analysis of precision, steadiness, and equilibrium in fine motor skills. *Amer. Psychologist*, 1950, **5**, 276-277. (Abstract)
21. Seashore, R. H. Work and motor performance. In S. S. Stevens (Ed.), *Handbook of experimental psychology*. New York: Wiley, 1951. Pp. 1341-1362.
22. Simonson, E., Enzer, N., & Blankstein, S. Effects of amphetamine (benzedrine) on fatigue of the central nervous system. *War Med., Chicago*, 1941, **1**, 690-695.
23. Simonson, E., & Enzer, N. The effect of Amphetamine (benzedrine) sulfate on the state of motor centers. *J. exp. Psychol.*, 1941, **29**, 517-523.
24. Somerville, W. The effect of Benzedrine on mental or physical fatigue in soldiers. *Canadian med. Ass. J.*, 1946, **55**, 470-476.
25. Thornton, G. R., Holck, H. G. O., & Smith, E. L. The effect of benzedrine and caffeine upon performance of certain psychomotor tests. *J. abnorm. soc. Psychol.*, 1939, **34**, 96-113.

(Accepted for publication June 5, 1953)

Vol. 67, No. 16 Whole No. 366, 1953

Psychological Monographs: General and Applied

Situational and Personality Factors in Leadership among Sorority Women[1]

Bernard M. Bass
Louisiana State University

Cecil R. Wurster
Human Resources Research Institute
Maxwell AFB, Alabama

Paddy Ann Doll
University of Texas

Dean J. Clair
Northwestern University

I. INTRODUCTION

The purpose of this study was to increase understanding of leadership among sorority women and to increase knowledge of the measurement potentialities of certain new or less well-known assessment procedures. Five kinds of measurements were collected, viz., biographical information, peer ratings, self-ratings, paper-and-pencil test scores, and situational test scores.

The factorial content of some of the specific measurements, such as American Council on Education Q and L scores, was well known; other measurements, such as age and year of education, were considered intrinsically valid. At the other extreme were measures whose content was unknown, such as attempts to assess in a systematic manner the thought processes of subjects by means of the "stimulated recall" procedure of Bloom (7). Thus, on the one hand, we had reference variables about whose factorial composition we were fairly sure; we also had variables such as the Guilford-Zimmerman Temperament Survey scores or leaderless group discussion (LGD) scores, about which we had some previous experimental evidence; and we had a mass of unknowns.

In addition to this exploration into the factorial composition of various measures, we also were in a position to examine the extent to which successful leadership behavior of individuals varied as a function of three situations: (*a*) individual sorority activities, (*b*) discussions among representatives from seven sororities, and (*c*) nonsorority university activities. In addition, study was made of the extent to which successful leadership behavior of individuals was a function of the mode measurement.

The reader will note our failure to

---

[1] Different phases of this study were carried out by different Louisiana State University graduate students working cooperatively under the direction of the senior author. Miss Paddy Ann Doll was responsible for collecting subjects and administering the personal histories, sociometric questionnaires, and paper-and-pencil tests; Mr. Dean Clair was responsible for the coding of the stimulated recall data; Messrs. Cecil Wurster and Robert McGehee were leaderless group discussion observers. The correlational matrix and factorial analysis were computed by Mr. Wurster with the aid of the IBM department at Louisiana State University. A number of others served on various phases of the project, including Messrs. William Hawkins, Arnold Gebel, and Stanley Klubeck. Financial aid was rendered by the Louisiana State University Graduate Council on Research.

The authors are indebted to Dr. Benjamin Bloom for aid in organizing the stimulated recall procedures.

mention the concepts of criterion, and of validity as evaluated by correlation with a criterion. We feel it would be unwise to label arbitrarily some measures as assessments or predictors and others as appraisals or criteria, since we were not particularly interested in the extent to which we could predict scores on some of the variables if scores on some others were known. McClelland (14) has summarized the rationale of this methodological position.

II. SUBJECTS AND ADMINISTRATIVE PHASES IN THE ASSESSMENT PROCESS

A. Subjects

The seven largest sororities on the Louisiana State University campus agreed to participate in this research project. The total membership of pledges and actives from each of the seven sororities was respectively 67, 53, 47, 46, 45, 41, and 36. Twenty volunteers were obtained from each of the seven sororities. These 140 girls were divided into 20 groups, each containing one representative from each sorority. Their scholastic standing ranged from freshman to senior in the university. The average subject was just under 19 years old (18.97) and had completed 2.16 semesters of college. (The respective standard deviations were .99 and 1.13.)

Before the beginning of the project, the seven sororities were told that the general nature and purpose of the program was to assess potential and actual leadership through psychological measures.

The subjects were motivated to volunteer in two ways: (*a*) each girl was promised that a personal letter would be sent to her discussing her assessment based on the various phases of the research in which she participated, and (*b*) each sorority would be informed as to its standing in relation to the other six sororities. This double incentive was successful in recruiting the necessary volunteers.

B. Administrative Phases

Each of the 20 groups of subjects was subjected to a three-phase testing program.

1. *First phase*. Each group of seven subjects, one subject representing each of the seven participating sororities, was convened for an initially leaderless group discussion of one-half hour. During this first phase, two observers rated the amount of successful leadership activity (LA)[2] each participant displayed during the discussion. Two other observers counted the number of responses each discussant made; this was used as an index of her participation (PN).

2. *Second phase*. The girls were ranked according to the extent to which each had displayed successful leadership behavior during Phase 1 and were given different assignments for the second phase according to their respective ranks attained. The girls ranked second, fourth, and fifth and as many of the girls of rank one, three, six, and seven as had the time, after completing other tasks (5, 12), were administered the Guilford-Zimmerman Temperament Survey. Forty-nine of the subjects were able to complete all ten subtests of the survey, while 76 completed at least the first five subtests.

All 140 subjects completed a series of

---

[2] Letters are the code assigned the measurement. See Table 1 for all the measurements and their code letters.

sociometric questionnaires concerning their own sorority. Peer ratings, NP, NO, NS, NR, VS, AD, and NI (to be described later), were obtained at this time.

Finally, all subjects filled out a personal history blank and a questionnaire concerning their motivation to be leaders. The measures (described later) which were obtained from this procedure included YR, CA, UA, UL, SL, and SS.

3. *Third phase.* The same group of seven as in the first phase was reconvened and subjected to a second initially leaderless group discussion from which the rating of amount of successful leadership activity (LB) exhibited by each participant was obtained.

The second discussion was recorded and immediately played back to the participants for the purposes of stimulated recall analysis to be described later.

4. *Other measures.* A number of other measures were obtained from records at the registrar's office. These included each subject's scholastic average (PA) and American Council on Education Psychological Examination scores (AQ and AL).

III. THE VARIABLES OF THE STUDY

We will now describe or define each of the variables, data on which were collected in this study.

The variables have been categorized according to the method by which scores for them were obtained. The five modes of data collection were: (*a*) biographical information blanks, (*b*) peer ratings, (*c*) self-ratings, (*d*) paper-and-pencil tests, and (*e*) situational tests. Table 1, a list of the variables and their codes, is provided for easy reference.

A. BIOGRAPHICAL INFORMATION

Biographical information was collected from school records and from a biographical information blank which all subjects filled out in the second phase between the first and second group discussions (or else, on another day). The seven variables are as follows:

1. *Semesters of education* (YR). The number of semesters of college work completed by a subject.

2. *Chronological age to the nearest birthday* (CA).

3. *Extracurricular, university, nonsorority activity per semester* (UA). Subjects were asked to list the names of all clubs, committees, and organizations they now belonged to or had belonged to on the LSU campus outside of the sorority. In addition, they were asked to indicate the number of semesters they had belonged to each group. The final index used was the total number of semesters of membership mentioned for all groups divided by the number of semesters on campus.

4. *University, nonsorority, leadership activity per semester* (UL). Subjects were asked to indicate the offices they had held or were holding in the organizations they had listed for determining UA. The more than one hundred offices mentioned were assigned weights by five judges from 0 to 4 according to their importance and significance in terms of the extent to which the office-holder of a designated position was a leader. The five judges were experienced student leaders not in the study. UL was obtained by assigning appropriate weights to each office reported by a subject, summing and dividing by the number of semesters the subject had spent on the campus. The last division was performed for the same reason as was given above for UA.

The UL was considered to be a rough estimate of the leadership status—as distinguishable from esteem—of the member at the university outside her sorority. University leadership status was conceived as the relative extent to which the university leadership positions held by a subject are of significance, importance, or value in contributing to the maintenance or progress of the group within which the position is found.

5. *Sorority leadership activity per semester* (SL). This index was derived in the same way as UL, except that the offices that subjects listed were inside instead of outside their sorority. SL was conceived as measuring the extent to which each subject had status at present and in the past in the sorority.

6. *Scholastic point average* (PA). This average or cumulative point-hour ratio was obtained for

each subject from the registrar. In the academic scoring system at LSU, A=3, B=2, C=1, and D and F=0. The sum of the cross products of grades earned for hours taken, divided by hours taken, yielded the point average.

The collapse of D and F into one category served to reduce the variance of this measure somewhat and thus correlations between PA and other measures tended to be somewhat conservative.

7. *Socioeconomic status* (SS). Each subject was asked, at a later date after the assessment had been completed, to indicate her father's occupation and education and also to indicate her own religion. Mentioned occupations were classified and assigned points as follows: professional, managerial, administrative, +3; white collar, sales, clerical, small business, +2; skilled laborer, +1; semiskilled and other, 0. Mentioned parental educational level was scored as follows: college graduate, +3; some college, +2; high school graduate, +1; and eighth grade, 0. Religion was a more difficult category to score because of the peculiarly different status of Catholics in south Louisiana, especially in the larger cities, as compared with most other parts of the United States. It was decided to score Episcopalians, Presbyterians, etc. as 2, Catholics and Methodists as 1, and Baptists and Church of Christ as 0. It is probable that this use of religion as a means of socioeconomic stratification was seriously in error and may have served to negate much of the usefulness of the socioeconomic index, which was the sum of the weights assigned for parental occupation, education, and religion. Furthermore, since the sorority subjects probably represented a severely restricted range in socioeconomic status, the likelihood was low that this index would provide much reliable discrimination.

TABLE 1

A LIST OF THE VARIABLES USED IN THIS STUDY AND THEIR CODE LETTERS

A. *Biographical Information*
    YR —Semesters of education
    CA —Chronological age to the nearest birthday
    UA —Extracurricular, university, nonsorority activity per semester
    UL —University leadership activity per semester
    SL —Sorority leadership activity per semester
    PA —Scholastic point average
    SS —Socioeconomic status
B. *Peer Ratings* (of Own Sorority Sisters)
    NP —Nomination for president
    NO —Nominations for organizer
    NS —Nominations for public speaker
    NR —Nominations for representative

TABLE 1 (*Continued*)

    VS —Visibility
    AD—Overt social adjustment
C. *Self-Ratings*
    NI —Self-esteem
    UM—Motivation to attain university leadership status
    SM —Motivation to attain sorority leadership status
D. *Paper-and-Pencil Test Scores*
    1. *American Council on Education Psychological Examination*
        AQ—Numerical aptitude
        AL —Verbal aptitude
    2. *Guilford-Zimmerman Temperament Survey*
        ZG —General activity, energy vs. inactivity, slowness
        ZR —Restraint, seriousness vs. impulsiveness, rhathymia
        ZA —Ascendance, social boldness vs. submissiveness
        ZS —Social interest, sociability vs. shyness, seclusiveness
        ZE —Emotional stability vs. emotional instability, depression
        ZO —Objectivity vs. subjectivity, hypersensitiveness
        ZF —Friendliness, agreeableness vs. hostility, belligerence
        ZT —Thoughtfulness, reflectiveness vs. thoughtlessness, extraversion
        ZP —Personal relations, cooperativeness vs. criticalness, intolerance
        ZM—Femininity vs. masculinity of emotions and interest
E. *Situational Test Scores*
    1. *Leaderless Group Discussions*
        LA —Amount of successful leadership activity displayed in the first initially leaderless group discussion
        LB —Amount of successful leadership activity displayed in the second initially leaderless group discussion
        PN—Amount of participation or attempted leadership activity in the first initially leaderless group discussion
        EF —Average efficiency of the two discussions in which subject participated
    2. *Stimulated Recall Analyses of Thought Processes*
        SE —Self-oriented
        AN—Oriented towards another person or persons
        GR—Oriented towards groups rather than individuals
        AP —Relevance to attaining formal group goals
        CE —Certainty or assuredness
        DA—Disagreement or agreement rather than neutral
        SN —Hostile and belligerent
        PR —Concerned with group processes and dynamics
        EV—Evaluative

Two raters independently scored the socioeconomic status for each subject according to the schedule described above. The correlation between their assigned values was .97.

B. PEER RATINGS (OF OWN SORORITY SISTERS)

Esteem has been defined elsewhere (4) as the extent to which a member of a group is perceived by the others in the group to contribute (or to be likely to contribute) to the maintenance and progress of the group—regardless of the member's position. Esteem is associated with the concept of meritorious performance.

On the other hand *status* can be regarded as the extent to which any member—regardless of his personal merit—can contribute to the maintenance and progress of his group because of the worth or importance of his position.

It has been shown both deductively and inductively that a member's potential to lead others successfully is a function of his status and his esteem. Therefore it was felt important to obtain estimates of the extent to which our subjects varied on both. Actual offices held (SL) was conceived to be a fairly "pure" measure of sorority *status*. Peer ratings were necessary to obtain adequate measures of sorority esteem.

It was assumed that if members were asked to choose each other for a position of importance to the sorority's maintenance and progress, such as president, the choice would be based primarily on expectations or perceptions of the worth of the possible nominees, *as persons*, to the sorority. Thus, it was expected that the nomination procedure would provide reasonable indices of esteem rather than status. However, it was realized that status and esteem, although conceptually distinguishable, would probably be correlated to some degree.

Since status is much more likely to be a matter of opportunity, seniority, fortuitous circumstances, and other factors unassociated with personal characteristics, it was hypothesized that esteem would provide a necessarily "purer" measure of leadership potential than status per se.

1. *Nominations for president* (NP). During the second phase, each of the subjects was presented a list of all the members of her own sorority, whether or not they were included in the 140 subjects of this study. She was asked to select the 21 (including herself) with whom she was most familiar. She then was requested to indicate the seven from the 21 she would most want to be president or hold some high office in the sorority, and the seven she would least want to be president or hold high office. Each rater thus was forced to rate only those girls with whom she was fairly familiar. It was hoped this would reduce the artifactual tendency of girls to be selected or rejected merely on the basis of visibility. Any remaining correlation between esteem and visibility would then have to be attributed mainly to elements other than measurement contamination.

Scores for each of the 335 members of the seven sororities were obtained in the following manner: Each ratee was credited with two points every time someone most wanted her for president, one point when in the selected 21 and in neither the top nor bottom seven in demand, and no points when chosen as least wanted for the presidency. Her total points were divided by the number of raters who placed her in the "21-most-familiar" category. Since no two members of the same sorority ever appeared together in a group discussion, peer ratings were free of contamination which might have occurred if any raters had had the opportunity of observing those they rated in the leaderless group discussion.

Table 2 shows the estimated reliabilities for NP for each of the seven sororities and the number of ratees per sorority. To obtain these reliability estimates, the twenty raters of a given sorority were split into two samples and ratees' NP scores based on ten raters were correlated with their scores based on the other ten and corrected by the Spearman-Brown formula. Table 2 shows that the estimated reliabilities for NP for each of the seven sororities were .85, .85, .78, .65, .90, .73, and .85 respectively. The median of .85 reflected the large number of raters upon whose judgment each rating score was based.

2. *Other nominations.* NO, NS, and NR were obtained in the same way as NP. All four were

concerned with the extent to which each subject was nominated by her sorority sisters for positions of significance to the sorority's existence. Hence, all four were conceived as relevant estimates of the sorority leadership potential of the subjects on the assumption of the validity of the hypothesis that individual differences in esteem in a group accurately forecast differences in future leadership activity.

NO concerned the nomination of subjects who "promote and organize the sorority's activities"; NS involved selection of members to address an audience of sorority alumnæ; NR concerned the choosing of peers to represent the sorority at a national convention.

Table 2 displays the corrected split-half reliabilities of NO, NS, and NR.

3. *Visibility* (VS). Visibility was the extent to which a subject was well known in her sorority. This measure was the number of times a ratee was placed with the 21 sorority members a rater was willing to nominate or reject during the voting to obtain NP, NO, NS, and NR. Reliability estimates were obtained in the same manner as for the preceding variables and are shown in Table 2. It was hypothesized that visibility would be associated with esteem. That is, a person would have to be well known in order to be selected for positions of importance.

4. *Overt social adjustment* (AD). During the second phase, a list of the members of her own sorority was given to each subject along with the following directions:

"Below is a list of all your sorority sisters. Strike out your own name and the names of all girls you do not know at all.

"Next, indicate which of the following descriptions generally applies to each name by placing the letter of the description which applies in front of each name. If you are in doubt about some letter assignment, place a circle around the letter or letters which you assign with uncertainty.

"Assign one letter to all the appropriate names before proceeding to assign the next letter."

Letter Code

A—idealistic, cooperative, adventurous, grateful, soft-hearted, friendly, enthusiastic, reasonable, trustful.

B—timid, withdrawn, thankless, hostile, secretive, irrational, pessimistic, unhappy, suspicious, short-tempered, obstructive

C—persevering, practical, self-controlled, calm, patient, loyal, honest, mature, deliberate

D—quitting, evasive, excitable, impatient, irritable, impulsive, unreflective, immature

E—boastful, conceited, assertive, headstrong, exhibitionist

F—slipshod, conscienceless, simple, undependable, absent-minded, slow, quitting, uninquiring

G—thoughtful, analytical, wide interests, conscientious, sophisticated, intelligent, independent mind

The number of B's, D's, E's, and F's assigned a ratee by those that rated her was subtracted from the A's, C's and G's and divided by the number of persons who rated her. The clusters represented seven surface traits obtained by Cattell (8) by factorial methods. It was assumed that these observable characteristics had at their roots a common source trait of social adjustment. The socially well-adjusted girl was thought to manifest A, C, and G traits while the socially maladjusted subject was thought more likely to exhibit B, D, E, and F characteristics.

C. SELF-RATINGS

Three self-ratings were obtained.

1. *Self-esteem* (NI). This measure was the sum of the +2 and +1 categories into which each subject assigned herself on NP, NO, NS, and NR. It was thought to provide an estimate

TABLE 2

CORRECTED SPLIT-HALF RELIABILITIES OF NP, NO, NS, NR, VS, NI, UM, AND SM, BY SORORITY

Variable	N	Sorority A	B	C	D	E	F	G	Mdn
NP	335	.85	.85	.78	.65	.90	.73	.85	.85
NO	335	.84	.94	.88	.85	.84	.77	.85	.84
NS	335	.83	.90	.83	.81	.73	.73	.60	.81
NR	335	.86	.74	.78	.74	.83	.77	.65	.77
VS	335	.85	.98	.74	.84	.63	.79	.89	.84
NI	140	.67	.80	.77	.71	.68	.68	.75	.71
UM	140	.78	.86	.84	.93	.94	.97	.80	.86
SM	140	.87	.59	.82	.93	.92	.91	.84	.87

of the subject's self-esteem—the extent to which she regarded herself as of worth to her sorority regardless of her present position. By comparing each subject's self-assignments on NP and NO with those on NS and NR, an estimate of the reliability of NI was obtained and corrected, sorority by sorority, (see Table 2). The median reliability was .71.

2. *Motivation to attain university leadership status* (UM). During the second phase, all 140 subjects were administered a questionnaire similar to the following:

"For each of the following university offices, please indicate your true feelings by placing next to the office the number of one of the five statements listed below that best describes it."

1. I would not serve even if I were drafted for the office.
2. I would serve only if drafted for this office.
3. I don't care whether or not I am elected or appointed to the office.
4. I should like to be elected or appointed to the office.
5. I would try my hardest to be elected or appointed to the office.

______Coed vice-president
______Student president of your college
______Any office in your college
______Student council representative
______President of your class
______High officer in any honorary organization
______High officer in a scholastic organization
______A.W.S. officer
______Dormitory officer
______Purple Jackets officer
______YWCA officer
______Officer in religious organization
______Interdormitory Council member
______Honor Court member
______Officer of intersorority group
______Officer in Panhellenic group
______Officer in professional club or group

The sum of all assigned numbers was conceived as an estimate of motivation.

Odd-even reliability corrected by the Spearman-Brown formula for each of the seven sororities is displayed in Table 2. The median reliability of .86 was considered satisfactory. It was hypothesized that many indices of leader behavior such as that exhibited during the leaderless discussions would be a function of motivation to be a campus leader.

3. *Motivation to attain sorority leadership status* (SM). This measure was obtained in the same manner as was UM. The sorority offices listed were as follows:

______President
______Vice-president
______Secretary
______Treasurer
______Pledgemistress
______Historian
______Scholastic chairman
______Membership chairman
______Panhellenic representative
______Sergeant-at-arms
______Magazine chairman
______Activities chairman
______Athletic chairman
______Parliamentarian
______Reporter, press chairman
______Social service chairman
______Social chairman
______Cultural program chairman

The obtained median reliability of .87 was considered adequate. (See Table 2.)

D. PAPER-AND-PENCIL TEST SCORES

Aptitudes and personality traits were assessed by well-established paper-and-pencil procedures.

The American Council on Education Psychological Examination provided two aptitude assessments and the Guilford-Zimmerman Temperament Survey provided ten personality measurements.

1. *Numerical aptitude* (AQ). This was the Q score in terms of local norms obtained from the ACE. These scores were obtained from a campus bureau responsible for testing entering freshmen or from appropriate agencies of other colleges from which subjects had transferred. ACE scores were collected for 136 subjects by these means.

2. *Verbal aptitude* (AL). This was the L score of the ACE. Both L and Q were considered to have satisfactory reliabilities and to be indicative of verbal and numerical aptitude. (For a survey of the literature on the ACE the reader is referred to Super [18].) These measures were included in this investigation because of the large number of studies available which suggested that the leaders in many kinds of social situations tend to surpass nonleaders in intellectual capacity. Moreover, the L score had been found correlated with leaderless group discussion behavior by Wurster and Bass (22) and by Bass and Coates in studies of leadership in ROTC.

3. *Guilford-Zimmerman Temperament Survey scores.* It was possible to obtain five subtest scores, ZG, ZR, ZA, ZS, and ZE, for 76 of the 140 subjects and all ten subtest scores for 49 on the Guilford-Zimmerman Temperament Survey. Unfortunately there was some bias in the selected samples. Those who scored highest or lowest in LGD score in the first of each of the

discussions (LA) were administered Rorschach tests or personal interviews during the second phase and were not generally available to take the Temperament Survey. Correlations between LA and the Temperament scores were thus attenuated slightly by the restriction in range in LA.[3] In addition it is expected that some slight attenuation may have appeared in the correlation between Temperament scores and any variables which correlated with LA as well as between Temperament scores which correlated with LA. No corrections for restriction in range were made, however, because the effects of the restriction were small and usually not direct.

The Guilford-Zimmerman Temperament Survey was selected to provide a battery of personality reference vectors for the factor analysis since the battery had been constructed by factor analytical methods. Some evidence was available concerning the battery's factorial and external validity. Adequate reliabilities were reported. Moreover, the inventory had been developed and standardized on a college population (11).

The definitions of the ten subtest scores are displayed in Table 1 and need not be repeated here.

It was recognized that the large amount of missing data in these tests would make correlations between them and other variables less reliable. The choice the investigators faced was eliminating them from consideration and losing these important means of connecting the results of our study to a large mass of personality literature, or taking a chance by assuming that the obtained correlations were reasonable estimates of what they would have been had there been no selective missing data. The investigators decided to risk being in error. However, the reader should keep in mind these imperfections when evaluating the results of the factor analysis.

E. SITUATIONAL TEST SCORES

1. *Leaderless Group Discussion*

Following is a description of the means by which each measure was obtained.

a. Amount of successful leadership activity displayed in the first initially leaderless discussion (LA). Each group of seven girls was subjected to two initially leaderless discussions, one in the first phase of the assessment program and one in the third phase. The seven girls each represented a different sorority and were relatively unacquainted with each other. This was done to make the leaderless discussion initially free of a status hierarchy. Bass and Wurster (6) have shown that leaderless discussion behavior is closely associated with any previous status structure which may exist among the participants. The extent to which the leaderless group discussion was free of such influence was demonstrated by the correlation of .10 which was found between sorority leadership status (SL) and leaderless discussion performance (LA). It was assumed that LGD performance would be associated primarily with permanent personality characteristics rather than the more situationally derived status. Thus, continued *successful* leadership activity displayed in the first leaderless group discussion could be attributed in large part to the esteem (perceived ability to contribute to the group's effectiveness) gained by each of the participants during the discussion. Esteem, in turn, would depend more on personal factors such as verbal ability and social boldness rather than on factors associated with participants' positions in their respective sororities, such as length of time in the sorority.

Since sorority status could undoubtedly also be associated with esteem, we would expect some correlation between sorority status and leaderless discussion success, but this would occur because both would be associated with some common personal factors such as social maturity.

The initially leaderless group discussion was administered as follows:

A designated group of seven girls was arranged in the shape of an inverted V with two well-trained observers seated at the open end. The subjects were given a list of questions such as "Should coeds marry while still in college?"; "Are girls' schools better than coeducational colleges?"; or "On what bases should sorority pledges be selected?" They were directed to select one question and to discuss it without help from the observers for 30 minutes to arrive at a set of conclusions about the question.

Fifty-two questions were randomly divided into four lists, only one of which was presented at each discussion. The questions for discussion were designed to be stimulating enough to provide for a full 30-minute discussion and to be of a general enough nature that it could be assumed that all participants had a fairly equal fund of knowledge about each issue. This helped keep at a minimum the effects of knowledge specific to a given situation as a source of individual differences in the attainment of leadership status.

LA was obtained by means of observers' ratings. The senior author has defined in detail elsewhere (4) the concept of leadership as used in this study. Leadership is conceived to be an interaction between any two or more members of a group. For two members, A and B, a *successful* leadership interaction takes place when B changes behavior in a direction desired by A as a consequence of A's act. If B fails to change, A's act is unsuccessful or *aborted*. For the group as a whole, the amount of successful leadership activity by one member, A, can be measured by the extent to which his behavior results in the movement of his group in a shared direction—that is, A's successful leadership is the extent to which the group is able and willing to change in a direction desired by A.

---

[3] The reader is referred to Bass, McGehee, *et al.* (5) for a more detailed account of this restriction and its effects.

Two trained observers independently indicated on a five-point scale the extent to which each participant of an initially leaderless group discussion displayed seven highly intercorrelated behaviors (av. $r = .85$), all of which were deemed items indicative of the extent to which each participant displayed successful leadership, i.e., moved her group in a shared direction. The seven items were:

1. She was effective in saying what she wanted to say.
2. She offered good solutions to the problem discussed.
3. She showed initiative.
4. She clearly defined the problem.
5. She motivated others to participate.
6. She influenced the participants.
7. She led the discussion (a summary rating).

A participant's LA score was the sum of points received for each of the seven items of behavior from both observers. The correlation between the two observers when corrected by the Spearman-Brown formula gave an estimated reliability of .91 for LA. The correlation of .75 between LA and LB, the same measurement made during the second discussion, was an estimate of both measures' test-retest reliabilities in this study.

b. Amount of successful leadership activity displayed in the second initially leaderless group discussion (LB). This was obtained during the second LGD in the same way as LA. LB had a corrected split-rater reliability of .84.

c. Amount of participation or attempted leadership activity in the first initially leaderless group discussion (PN). Two observers, other than those whose estimates were used to obtain LA, recorded the number of responses emitted by each discussion participant. PN is the simple sum of all these recorded responses. PN was included in the study because it provided a means for testing a number of hypotheses concerning attempted, successful, and aborted leadership behavior. A number of earlier studies (1, 2) have disclosed the existence of correlations ranging from .65 to .91 between various indices of participation and tendency to succeed as a leader.

It was assumed that almost all participation in a highly ambiguous social situation could be considered attempts to change the perception or behavior of others, hence almost all participation in such situations could be considered *attempted* leadership.

Attempted and successful leadership activity correlate by definition (4), for all successful activity must be attempted. But the reverse does not hold. Attempted leadership activity may be either successful or aborted. Thus, the part-whole correlation between the two may be regarded as expressing a nominal rather than a postulational or empirical relationship.

The mean number of responses per subject was 149.6. The correlation between the two observers gave an estimated split-half reliability of .95 for PN.

d. Discussion efficiency. Two observers rated the extent to which each of the 20 pairs of discussions was efficient—i.e., attained the goals set forth in the problem selected by the group with a minimum waste of time and effort. The average correlation between the two observers' five-point ratings of efficiency for the 20 first-phase discussions and second-phase discussions was .75. The mean efficiency of the two discussions in which a subject participated was assigned to her as a discussion efficiency score. It was hypothesized that variations in discussion efficiency would be associated with various measures of individual differences.

2. *Stimulated Recall Analyses of Thought Processes*

a. Method. The stimulated recall analyses were based on a technique developed by Bloom (7). The second initially leaderless discussion was recorded on tape. As the discussion progressed, the senior investigator made note of ten "critical points" along the way (e.g., a long silence occurred, a dogmatic assertion was made, a question was posed). The tape recording of the discussion was played back immediately to the girls. Each time the recording was stopped at a critical point, the subjects wrote down what they believed their thoughts were at that point during the discussion.

b. Content analysis. Content analyses of the protocols of each subject were made by two trained raters. The mode of analysis was "tailor-made" for the present study, although many of the categories had been used earlier by Bloom.

Since productivity varied considerably because a subject could report more than one thought per critical point (and many did), and since productivity per se was not of prime concern in the analysis, the number of thoughts coded in a particular manner was converted to a percentage of the total number of thoughts expressed. Each thought was classified into a number of different categories which were not necessarily mutually exclusive.

c. Indices. Listed below are each of the final indices which were obtained from the analysis. Estimates of the reliability are either based on the correlation (corrected by the Spearman-Brown formula) between two raters' independent categorizations of a random sample of 70 of the 140 protocols, or are mean estimates.

1. Self-oriented thoughts (SE), $r=.82$.
2. Thoughts oriented toward another person or persons (not groups) (AN), $r=.76$.
3. Thoughts oriented toward groups rather than individuals (GR), $r=.82$.
4. Thoughts relevant to attaining formal group goals (AP) $r=.56$.
5. Thoughts certain or dogmatic (CE), $r=.67$.
6. Thoughts approving or disapproving rather than neutral (DA), $r=.64$.
7. Thoughts hostile or belligerent (SN), $r=.68$.
8. Thoughts concerning group processes, group dynamics, interaction among persons (PR), $r=.50$.
9. Thoughts evaluative (EV), $r=.50$.

IV. RESULTS

Matrix of intercorrelations. Pearson product-moment correlations were computed among all 41 variables. The resulting matrix of correlations among all except the stimulated recall variables is displayed in Table 3. Table 4 shows the

TABLE 3

INTERCORRELATIONS AMONG THE ASSESSMENT VARIABLES

Variable	Biographical Information							Peer-Ratings						Self-Ratings		
	YR	CA	UA	UL	SL	PA	SS	NP	NO	NS	NR	VS	AD	NI	UM	SM
YR		74*	60	43	41	08	03	23	28	23	27	19	21	03	−04	−23
CA			53	41	33	11	00	20	27	23	24	16	25	10	−02	−17
UA				62	36	18	02	28	37	33	34	30	27	18	03	−16
UL					23	27	−12	35	35	43	38	23	29	18	04	−17
SL						22	−01	48	55	49	55	44	25	20	09	−12
PA							10	29	24	32	22	16	33	14	−06	−15
SS								00	−03	−09	−13	−03	00	−09	−06	−04
NP									77	80	82	48	65	30	15	00
NO										76	78	59	53	39	07	−06
NS											90	57	61	40	17	−01
NR												58	61	34	14	00
VS													34	25	00	−19
AD														17	16	−03
NI															21	13
UM																55
SM																
AQ																
AL																
ZG																
ZR																
ZA																
ZS																
ZE																
ZO																
ZF																
ZT																
ZP																
ZM																
LA																
LB																
PN																
EF																

---

* Correlations in boldface are significant at the 1 per cent level of confidence. Decimal points have been omitted.

intercorrelations among the stimulated recall variables. To save space, the correlations between the stimulated recall variables and the others have been omitted since all but two were not significant at the 1 per cent level of confidence.

The factor analysis. The matrix of intercorrelations was factored by means of a group clustering method of multiple-factor analysis outlined by Cattell (9). The centroid axes were rotated to psychologically meaningful positions and to as simple a structure as possible. Orthogonality was maintained between all rotated factors for reasons which will be cited later, although an oblique solution for the first two factors would have been superior.

Table 5 shows the 11 factors which remained after rotation. The loadings of variables on the eleventh factor were so small upon completion of rotation that this last factor was considered to account for a negligible amount of the common variance among the measurements.

The communality of each assessment is shown, together with the variance specific to the variable where it was possible to obtain reliability estimates from the data.

TABLE 3 (*continued*)

(EXCEPT THE STIMULATED RECALL INDICES)

ACE Scores		Guilford-Zimmerman Temperament Survey Scores										Situational Test Scores			
AQ	AL	ZG	ZR	ZA	ZS	ZE	ZO	ZF	ZT	ZP	ZM	LA	LB	PN	EF
−06	06	09	11	−21	−08	10	19	14	−16	10	11	18	11	12	27
−07	03	12	10	−18	−06	08	13	16	02	03	02	20	13	12	27
13	04	17	07	02	14	08	−01	−03	−07	25	01	32	21	27	27
15	08	27	02	27	16	21	30	15	−10	18	18	36	25	27	28
07	−08	11	15	−29	06	20	14	09	−11	−03	16	10	10	07	20
47	33	−01	33	10	−22	04	00	−08	40	06	03	31	25	26	07
−02	21	03	22	00	02	29	08	08	−03	26	26	03	−04	−08	06
13	−03	20	12	−05	02	08	25	29	00	24	14	21	14	10	08
05	02	12	06	16	21	03	08	12	−13	22	17	33	25	19	12
18	02	−01	02	04	04	11	18	25	−18	15	09	36	27	21	12
15	09	09	10	−08	07	09	20	21	−20	21	14	27	21	18	10
09	−05	−01	02	10	17	09	−14	−01	−08	12	−08	24	18	14	14
12	−02	−04	12	−18	−24	−01	41	39	02	15	−04	28	19	11	19
04	−07	−06	−05	14	30	14	−04	01	−02	22	26	29	31	21	15
−14	−20	−15	−11	−01	−10	−12	15	−02	−02	11	18	−03	00	−11	04
−22	−10	−07	−19	16	−06	−04	25	−05	−06	27	22	−07	−05	−10	−06
	45	00	27	−09	−34	07	−19	−12	36	−19	09	17	18	21	06
		−14	26	04	−14	14	07	−02	19	05	03	35	31	21	01
			−02	40	49	28	−07	−13	−09	−13	08	13	23	21	−03
				−19	−15	11	05	20	34	−23	−21	16	09	00	07
					68	34	−06	−26	−08	−01	−04	37	42	35	04
						27	−04	−17	−14	−01	−03	22	14	17	00
							45	29	−11	−01	25	16	15	05	00
								57	−27	30	38	−04	−09	−31	06
									−07	05	16	−16	−16	−12	−05
										−24	−15	13	03	−09	−15
											−08	10	−11	−10	30
												02	−05	01	−56
													75	77	12
														60	12
															−03

TABLE 4

INTERCORRELATIONS AMONG THE STIMULATED RECALL VARIABLES

	SE	AN	GR	AP	CE	DA	SN	PR	EV
SE		06	−01	**38***	**29**	**25**	−05	−06	**39**
AN			−09	**−35**	**−30**	02	**45**	**24**	−16
GR				12	**21**	**25**	**−25**	**−73**	−07
AP					**46**	**46**	−34	−12	**48**
CE						**54**	−05	**−28**	**36**
DA							08	−13	**36**
SN								**33**	00
PR									11
EV									

* Correlations in boldface are significant at the 1 per cent level. Decimal points have been omitted.

TABLE 5

FINAL ROTATED ORTHOGONAL FACTOR MATRIX

(Decimal points have been omitted.)

Variable	Factor											h^2	s^2	r_{11}
	I	II	III	IV	V	VI	VII	VIII	IX	X	XI			
YR	24	68	−07	−05	02	08	03	−07	−01	00	01	54		
CA	22	65	−01	−01	04	03	03	−04	−09	03	−03	49		
UA	31	57	12	−06	04	−03	09	−02	−09	−05	−03	46		
UL	32	44	25	01	07	19	05	00	−03	03	07	41		
SL	52	28	−06	01	−01	06	02	−01	01	06	−04	36		
PA	26	06	04	05	42	−07	17	−04	03	03	03	29		
SS	−05	02	01	−05	07	21	02	−03	11	03	−14	09	89	98
NP	73	12	03	05	01	12	−02	−01	04	06	−02	57	28	85
NO	72	15	17	04	01	01	04	01	05	08	00	58	27	85
NS	76	11	05	01	03	10	10	−07	01	03	01	62	19	81
NR	82	11	04	01	03	09	03	13	02	01	−01	71	19	90
VS	54	09	19	02	−02	−07	05	09	−10	−02	01	36	48	89
AD	54	11	−11	05	06	14	02	01	−07	00	00	35		
NI	31	02	18	02	−03	01	14	17	−23	03	−02	23	48	71
UM	12	−02	−11	00	−15	−08	−01	48	06	−03	03	29	49	78
SM	−02	−18	−01	03	−15	−07	00	48	13	−14	00	33	54	87
AQ	10	07	05	00	43	−09	08	−12	03	05	−06	24		
AL	−01	05	04	03	36	09	29	−09	01	06	−08	25		
ZG	07	14	43	−05	−10	02	03	−08	−01	00	05	23		
ZR	09	07	−04	07	28	04	05	−11	−10	06	−09	14		
ZA	−06	−10	65	07	−02	02	13	11	−05	23	06	53		
ZS	05	05	57	−10	−26	04	03	−08	08	−07	05	43		
ZE	09	09	29	05	06	36	−01	−04	18	−13	14	31		
ZO	11	16	−10	−03	03	59	02	−22	10	02	−02	46		
ZF	21	06	−21	−04	−05	54	02	−02	02	−03	−04	39		
ZT	−02	−08	−08	02	31	−13	06	−03	00	09	−04	14		
ZP	16	10	−08	−08	−14	18	14	20	01	−06	05	17		
ZM	11	06	00	23	03	33	00	−19	48	−06	13	47		
LA	24	18	39	−04	21	01	51	02	−03	01	01	55	20	75
LB	18	10	40	01	19	01	36	−03	−03	02	05	37	38	75
PN	12	14	39	−08	16	01	45	−05	03	06	00	43	52	95
SE	09	−18	−01	40	08	−01	−10	−04	02	02	05	22	60	82
AN	03	13	03	00	06	−08	05	00	−19	31	−24	22	54	76
GR	−04	01	25	07	−19	07	−02	01	−01	59	03	46	36	82
AP	00	−09	−09	54	−02	03	00	01	06	15	−19	37	19	56
CE	−05	01	−05	46	−05	−04	−06	−01	−08	−11	23	30	37	67
DA	07	07	−08	50	−03	06	−07	−01	05	−21	−06	33	31	64
SN	06	02	03	08	08	−11	−07	−08	03	31	−29	22	46	68
PR	09	02	−02	−11	20	−06	06	−08	01	64	−04	49		
EV	04	03	−04	50	16	−02	04	−01	16	06	01	31		
EF	11	28	00	05	08	−12	−03	00	−48	03	04	35	40	75

V. CONCLUSIONS

A. THE FACTORS

We will first describe each of the 11 factors which emerged from the analysis in terms of variables with highest loadings on the respective factors. Then we will consider the factorial composition of each of the variables of central interest. Wherever possible we will attempt to indicate the practical or theoretical significance of the results as they are presented.

1. *Factor I: Sorority Leadership Potential*

Seven variables had loadings above .50 on Factor I. Four (NP, NO, NS, NR) were measures of sorority esteem, two (VS, AD) were closely associated with esteem, and one (SL) was a measure of present and past status in the sorority. Much inductive and some deductive support exists to maintain the validity of the two hypotheses: (*a*) the higher one's esteem, the higher will be one's future leadership activity, and (*b*) the higher one's status, the higher will be one's future leadership activity (4). Thus, all seven variables had in common an association with future sorority leadership activity. Sorority Leadership Potential was selected to denote the common underlying factor which differentiated subjects in their future leadership activity.

Variable	*Loading on Factor I*
NR—Nominations for sorority representative	.82
NS—Nominations for public speaker	.76
NP—Nominations for president	.73
NO—Nominations for organizer	.72
VS—Visibility in sorority	.54
AD—Overt social adjustment	.54
SL—Sorority leadership activity per semester	.54

This result corroborated Wherry and Fryer's isolation of a "pure" leadership factor based on buddy ratings (19) since it will be noted that all measures but SL were peer ratings. It is fortunate that our status measure, actual sorority leadership activity per semester, SL, provided a means of "anchoring" all the other variables; otherwise we would have cause to fear that Factor I was merely a measure of halo, reputation, or popularity, without reference to leadership potential.

2. *Factor II: University Experience and Activity*

Four variables defined this factor, which appears to be determined primarily by the length of time a subject had been at the university and the extent to which a subject engaged in extracurricular activities and leadership activities outside the sorority.

Variable	*Loading on Factor II*
YR—Semesters of education	.68
CA—Age to the nearest birthday	.65
UA—Extracurricular university activity per semester	.57
UL—University leadership activity per semester	.44

It is interesting to note that the loadings of the defining variables of Factors I and II could have been raised appreciably by rotating to an oblique solution in which there would have been a correlation of .57 between Factors I and II and where the factors would have remained the same in composition. This oblique solution was not used since an orthogonal structure obtained among all the other factors and the oblique solution would not have materially altered any of the factor loadings of the other variables.

3. *Factor III: Ascendance and Sociability*

Two variables, ZA and ZS, defined Factor III although six variables correlated .39 or higher with the factor. The factor appeared to resemble The Ohio State Leadership Studies factor "initiation of structure" (10, 12). The reader should note that these loadings would undoubtedly have been higher had the sample not been restricted to sorority women.

Variable	*Loading on Factor III*
ZA—Ascendance & social boldness vs. submissiveness	.65
ZS—Social interest & sociability vs. shyness & seclusiveness	.57
ZG—General activity & energy vs. inactivity & slowness	.43
LB—Amount of successful leadership activity in the second leaderless group discussion	.40
LA—Amount of successful leadership activity in the first leaderless group discussion	.39
PN—Amount of participation or attempted leadership activity	.39

4. *Factor IV: Critical, Relevant Thinking*

This was the first "stimulated-recall" factor to emerge. It appears to involve the extent to which subjects tended to think critically and with relevance to the issues at hand. Women who did so appeared at the same time to be somewhat self-oriented, yet sure of themselves in their thoughts. Factor IV was independent of Group Process Thinking, Factor X.

Variable	*Loading on Factor IV*
AP—Thoughts relevant to attaining formal group goals	.54
DA—Disagreement or agreement rather than neutral	.50
EV—Evaluative thinking	.50
CE—Certain rather than doubtful thinking	.46
SE—Self-oriented thinking	.40

5. *Factor V: Intellectualism*

It is necessary to consider loadings as low as .31 in order to define Factor V, which appears to involve scholarship, intelligence, and the tendency to be thoughtful and reflective.

Variable	*Loading on Factor V*
PA—Scholastic point average	.42
AQ—Numerical aptitude	.43
AL—Verbal aptitude	.36
ZT—Thoughtfulness & reflectiveness vs. thoughtlessness & extraversion	.31

6. *Factor VI: Freedom from Neuroticism*

This factor involved the lack of tendency of subjects to be emotionally unstable, hypersensitive, and hostile according to the results of the Guilford-Zimmerman inventory.

Variable	*Loading on Factor VI*
ZO—Objectivity vs. subjectivity & hypersensitiveness	.59
ZF—Friendliness, agreeableness vs. hostility, belligerence	.54
ZE—Emotional stability vs. emotional instability, depression	.36

7. *Factor VII: Verbality*

This factor appears to concern primarily variations among subjects in discussion participation or the tendency to attempt leadership activity in initially leaderless group discussions (PN); to be successful as a discussion leader (LA, LB); and to be high in verbal aptitude (AL).

The naming of this factor was partly based on the lack of correlation of non-discussion leadership or ascendancy variables or of discussion efficiency with it. The part-whole nature of the correlation between LA and PN was kept in mind in arriving at the meaning of this factor; that is, to be successful as a leader, one must attempt leadership acts, or talk. Conversely, to be permitted and to desire to continue to talk, one must exhibit some success as a leader in this situation. Another name for this factor might have been "discussion participation."

Variable	*Loading on Factor VII*
LA—Amount of successful leadership activity in the first leaderless group discussion	.51
PN—Amount of participation or attempted leadership activity	.45

LB—Amount of successful leadership activity in the second leaderless group discussion	.36
AL—Verbal aptitude	.29

8. *Factor VIII: Motivation to Increase Leadership Status*

This factor was a "doublet," based on the correlation between UM and SM.

Variable	*Loading on Factor VIII*
UM—Motivation to attain university leadership status	.48
SM—Motivation to attain sorority leadership status	.48

9. *Factor IX: Feminine Inefficiency*

This "doublet" was somewhat of a surprise. It appears to confirm cultural stereotypes about feminine ineffectiveness. Discussions among highly feminine women tended to be less effective than those among women more masculine in attitudes and in interests.

Variable	*Loading on Factor IX*
ZM—Femininity	.48
EF—Discussion effectiveness	—.48

10. *Factor X: Group-Process Thinking*

This "doublet" concerned variations in the tendency of discussion participants to focus their thoughts on the group process, or the group dynamics.

Variable	*Loading on Factor X*
PR—Thoughts concerned with group processes and dynamics	.64
GR—Thoughts oriented toward groups	.59

11. *Factor XI: Withdrawal-from-the-Field Thinking*

The last factor, XI, appeared to have very slight loadings on three stimulated-recall variables; if these are not mere chance relations, a factor may exist which denotes a tendency to "withdraw from the field." The loadings are too low to be anything but suggestive.

12. *Summary of Loadings*

In summary, the 11 factors which emerged in this analysis which accounted for most of the common variance in the basic 41 variables of the assessment program were:

I. Sorority Leadership Potential
II. University Experience and Activity
III. Ascendance and Sociability
IV. Critical, Relevant Thinking
V. Intellectualism
VI. Freedom from Neuroticism
VII. Verbality
VIII. Motivation to Attain Leadership Status
IX. Feminine Inefficiency
X. Group-Process Thinking
XI. Withdrawal-from-the-Field Thinking

The relatively large amounts of specific variance remaining following the extraction of 11 factors was due primarily to the fact that the assessment program included a fairly wide assortment of measurements from real-life situations, quasi-real situations, and paper-and-pencil tests.

To some extent, "measurement factors" emerged. That is, some factors tended to contain a cluster of variables based on the same type of measurement, namely, Factors I, II, IV, VI, VIII, and X. Moreover, the sample was highly restricted. Therefore, the correlations within the matrix were probably significantly attenuated, reducing the size of the communalities, h^2. The size of the communalities which would be found for the same variables among a general population of late adolescents of both sexes and more varied education would probably be higher.

B. FACTORIAL COMPOSITION AND CORRELATES OF LEADERSHIP AND ASSOCIATED VARIABLES

1. *Biographical Information*

a. Semesters of education and *age* were correlated sufficiently highly with each other ($r = .74$) to be considered simultaneously. Both were originally included because it was thought that differences might emerge in leadership behavior which could be attributed to scholastic acceleration or retardation of the subjects. Such differences did not appear.

As displayed in Table 3, both variables determined the extent to which subjects were active in extracurricular activities ($r = .60, .53$), were active as university leaders ($r = .43, .41$), and as sorority leaders ($r = .41, .33$). Also the esteem of the subjects depended somewhat on age and semesters of education (r's ranged from .20 to .28 with NP, NO, NS, and NR). Finally, motivation to be a sorority leader tended to decline with semesters of education and age ($r = -.23, -.17$), although motivation to be a university leader remained constant as the women matured.

Between two-fifths and half the variance of these two variables were associated with Factor II, University Experience and Activity, while a small portion of the remaining common variance was associated with Factor I, Sorority Leadership Potential.

b. Extracurricular, university, nonsorority activity per semester (UA). While 32 per cent of the variance in extracurricular activity at the university was associated with Factor II, University Experience and Activity, another 10 per cent was associated with Sorority Leadership Potential. The fairly high correlations between age, semesters of education, and activity probably occurred because the new student has to devote time and energy in adjusting to college life and in learning the opportunities available before being able to engage freely in extracurricular activities. Also, the development of interests in particular activities may require further maturation.

The correlations of .62 with university leadership activity, .36 with sorority leadership activity, and the factor loading of .31 on Sorority Leadership Potential tended to support one of Stogdill's (17) generalizations that over a wide range of situations, leaders tend to participate more in activities than nonleaders. More specifically, the relation between Sorority Leadership Potential and university participation corroborated a similar finding by Williamson for Minnesota students (20).

Sheer experience in university activities appears to make subjects more active and effective as LGD participants, since UA correlated significantly with all measured aspects of discussion performance (LA, LB, PN, EF). The common factors of Verbality, Ascendancy and Sociability, and Feminine Inefficiency did not account for these correlations, since UA was unrelated to assessments outside the discussion situation. Discussion performance per se appears associated with past activity in nonsorority university groups.

c. University leadership activity per semester (UL), appeared to be associated somewhat more with University Experience and Activity ($r = .44$) than with Sorority Leadership Potential ($r = .32$).

Similarly to a study by Williamson and Hoyt (21) which suggested that personality differences were more likely to distinguish student university leaders from nonleaders than sorority or fraternity leaders from nonleaders, we found that while students high in Ascendance and Sociability and in Freedom from Neuroticism tended to engage in university leadership activity, no personality factors discriminated sorority leaders from nonleaders as well.

In the same way as UA, university leadership was significantly associated with all measured aspects of leaderless group discussion performance per se.

d. Sorority leadership activity per semester (SL) was primarily a function of Factor I, Sorority Leadership Potential, and, to a lesser extent, of Factor II, University Activity and Experience. It appeared to be independent of measured situational and personality test factors. Whether a member actually held various positions of leadership in her sorority appeared unrelated to her Ascendance and Sociability, Verbality, Freedom from Neuroticism, etc. As was suggested before, sorority status, as such, was more dependent on temporary, situational, or nonpersonal factors such as years of education and age, rather than on consistent personal characteristics.

We suggest that compared with sorority status, university status was more associated with personality characteristics because esteem in the university group—the extent to which a woman as a person could contribute to the university group—was a more important determinant of her election to university positions of worth than was her corresponding sorority esteem in her election to positions of worth in the sorority. (As will be noted later, sorority positions, as a whole, may be of lower worth to the members than university positions. There may be less concern over who obtains sorority positions on the LSU campus.)

The evidence to support the validity of the hypothesis is slight but in the right direction. We had no measure of campus esteem corresponding with our indices of sorority esteem. However, we did have campus measures of ability which could be assumed to be associated with variations in esteem. While point average (PA) correlated .27 with UL, it correlated only .22 with SL. Numerical aptitude (AQ) correlated .15 with UL and .07 with SL. Verbal aptitude (AL) correlated .08 with UL and −.08 with SL. Success and activity in leaderless discussions (LA, LB, PN, EF) correlated significantly with UL but not SL.

The specific correlation of .23 between UL and SL tended to support the conclusion that some consistent differences in tendency to engage in leadership activities exist across different situations, although the consistency is much less than would be suggested in the popular literature on leadership.

While university leaders reported themselves as being somewhat socially bold ($r = .27$), sorority leaders reported themselves as submissive ($r = -.29$). If these correlations are not a result of random fluctuation, we may account for the results by arguing that university-wide women leaders have to exhibit some aggressive behavior, especially when competing with male students for various leadership offices. Within the sorority, the ideal or norm for women, especially Southern women, is overt social submissiveness. This cultural ideal may exert an influence on the personality required of any woman who would be a sorority leader. Outwardly, at least, social boldness must be repressed by the aspiring sorority officer if she is to be popular, accepted, and elected by her sorority sisters to positions of leadership.

e. The correlations of .27 and .22 between *scholastic point average* (PA) and university leadership and sorority leadership activities, and the correlations of .29, .24, .32, and .22, respectively, with the esteem ratings, NP, NO, NS, and NR, are in line with expectations and Stogdill's (17) conclusion that leaders tend to be higher in achievement than nonleaders. It was interesting to note, however, that overt social adjustment correlated .33 with academic achievement—contrary to popular notions about the social inadequacies of the good student.

f. *Socioeconomic status,* as measured in this analysis, appeared to be of little importance in increasing our understanding of leadership.

2. *Peer Ratings (of Own Sorority Sisters)*

a. The first four peer ratings (NP, ND, NS, and NR) were similar in factorial composition. All had high loadings on Factor I, Sorority Leadership Potential, and all had slight loadings on University Experience and Activity. Some differences exist, most of which appear reasonable. The purest measure of Sorority Leadership Potential was the nomination for a position as a representative of the sorority (NR). It had a loading of .82 on

Factor I. The least pure of the four was the nomination as organizer (NO), which had a loading of .72 on Factor I and a much lower loading of .19 on Factor III, Ascendance and Sociability. The relatively low loading of NO on Sorority Leadership Potential was probably a function of the cultural ideal of feminine social submissiveness described earlier.

These various measures of esteem were more highly correlated with various measures of ability and personality than was the measure of sorority leadership status (SL). However, the differences were not so great as had been expected.

b. Visibility in the sorority (VS) appeared to be associated with Sorority Leadership Potential ($r = .54$), and to a lesser extent with Ascendance and Sociability ($r = .19$). Even in this setting, with the cultural restriction placed on the leader's boldness, it appeared fairly obvious that persons high in leadership potential—potential to attain leadership status—are more familiar, better known, and more highly *visible* to their sorority sisters than persons low in each potential. In this, as in many other situations, the potential leader must and does actively participate in various activities which make her well known to others.

Elsewhere (4) the senior author has suggested that status and visibility would be more highly associated than esteem and visibility. The present study did not support the validity of this proposition, for visibility was slightly less correlated with SL than with NP, NO, NS, and NR. Possibly, some measurement contamination was responsible for this outcome.

c. Overt social adjustment, popularity, or *being well thought of socially in the sorority* (AD) was primarily related, as expected, to Sorority Leadership Potential ($r = .54$). It had less than expected association with Freedom from Neuroticism ($r = .14$).

For reasons the reader is now familiar with, there appeared to be a slight tendency for socially bold girls to be considered somewhat maladjusted ($r = -.18$). Girls with high social interest (ZS) were likewise considered so ($r = -.24$).

It appears that the sorority leader faces a peculiar role conflict. To lead, she must exhibit initiative. Yet, the exhibition of too much initiative in sororities is out of line with the norms for Southern women. Further exploration of how sorority leaders solve this particular conflict may provide a fruitful example of the more general conflict that most leaders face. As Hemphill (12) and Smith (16), among others, have pointed out, leaders usually offer to their followers both promises of reward and threats of requiring excess energy expenditure, loss of individualism, and so forth. The adequacy of a leader may depend primarily on his resolution of this role conflict.

Others in the sorority may be in an even greater dilemma. Thus, the seriousness of this conflict to the girl with initiative but without the esteem needed to be sorority leader is summarized by the following hypothesis:

A socially bold sorority member, a member with a great deal of initiative and ascendancy, who lacks esteem will be highly maladjusted socially.

To test this hypothesis an index was derived based on the difference between a subject's standard score on social boldness (ZA) and her standard score on the sum of the four peer ratings of esteem, NP, NO, NS, and NR. This index correlated $-.68$ with overt social adjustment (AD), strongly confirming our expectation. The index also correlated $-.36$ and $-.16$ with ZF and ZO, friendliness and objectivity, the two best measures of Freedom from Neuroticism.

3. *Self-Ratings*

a. Self-esteem (NI), an important determiner of leadership behavior (4), appeared to be factorially complex. Only 10 per cent of its variance was associated with actual Sorority Leadership Potential. Five per cent was associated with Feminine Inefficiency, 3 per cent with Ascendance and Sociability, 3 per cent with Motivation to Attain Leadership Status and 2 per cent with Verbality. The results appear to comprise a consistent picture which suggests that self-esteem may represent a compromise between beliefs about one's actual leadership potential and desired leadership potential.

b. Motivation to attain university leadership status (UM) was significantly related to motivation to increase sorority status ($r = .55$) and self-esteem ($r = .21$). Either motivation was fairly independent of many of the other assessment variables, or else the index used was not a very relevant measure of motivation. We choose the latter, more conservative interpretation. It is possible that a more indirect or projective procedure would have been useful. The high reliability of UM and SM suggests that something was measured consistently—what it was, was not revealed by the factor analysis.

c. Motivation to attain sorority leadership status (SM) correlated significantly with lack of university experience ($r = -.23$) and lack of numerical aptitude ($r = -.22$), in addition to its association with motivation to be a university leader. It also tended to be more negatively associated with various indices of leadership status, actual and potential, than was UM.

More value was attached by these subjects to university nonsorority offices than to sorority positions, for the mean index of desire for sorority office holding was 3.3 while it was 3.7 for holding campus positions. The difference (almost a whole standard deviation) was significant at the 1 per cent level of confidence.

The factorial composition of SM and the percentage of variance in SM accounted for by each factor was as follows: Motivation to Attain Leadership Status, 23%; Lack of University Experience and Activity, 3%; Lack of Ascendance and Sociability, 2%; Feminine Inefficiency, 2%; and Lack of Group-Process Thinking, 2%.

4. *Paper-and-Pencil Test Scores*

Since the ACE and Guilford-Zimmerman Temperament Survey scores were included as reference tests and were not of central interest to the study, they will not be considered here.

5. *Situational Tests*

a. The factorial composition of the amount of successful leadership activity displayed during the first discussion (LA) turned out to be surprisingly complex. At the same time, all but 20 per cent of the consistent variance of LA was accounted for.

Twenty-six per cent of the variance in LA was associated with Verbality. This confirmed a number of other studies which suggested the importance to discussion leaders of a verbal productivity or verbal fluency factor (22). Specifically, as in an earlier study by Wurster and Bass (22), LA correlated more highly with verbal aptitude ($r = .35$) than with numerical aptitude ($r = .17$). The correlation of .77 with discussion participation was in line with all studies on the subject (1, 2, 3, 15); as explained earlier, this correlation is partly a function of definition. The highest correlation of LA with peer ratings was with NS—nominations for public speaker ($r = .36$).

Fifteen per cent of the variance in LA was associated with Factor III, Ascendance and Sociability. The correlation of .37 with ZA itself suggested that social boldness was the common element involved. Moreover, we note that the next-to-highest correlation between LA and the peer ratings was with NO, nominations for organizer. These results tend to confirm an analysis of ROTC cadets by Bass and Coates which indicated that LA tended to forecast peer ratings of tendency to initiate structure and interaction but not the tendency to be considerate of subordinates—two of the independent leader-behavior factors isolated by Fleischman (10).

Six per cent of the variance was associated with Sorority Leadership Potential. Keeping in mind our discussions of the cultural norm of social submissiveness for Southern women—especially in the sorority situation—we suggest that these results were probably lower than would have been found for male subjects. While LA was not of much use by itself as a forecaster of Sorority Leadership Potential among Southern sorority women, it correlated .36 with UL, university leadership activity (in contrast to .10 with sorority leadership activity), for as pointed out earlier, ascendancy of women is probably more acceptable on campus than in the sorority.

Four per cent of the variance of LA was a function of Intellectualism and 3 per cent was a function of University Experience and Activity. This last result suggests that LA is related slightly to mere exposure to group activity and experience per se in groups.

b. Although they have no obvious practical application, the results based on the amount of leadership activity displayed in the second discussion (LB) may have theoretical significance. There was a doubling in the variance unique to discussion success in the second discussion and an equivalent loss in variance associated with the factors of Verbality, Sorority Leadership Potential, and University Experience and Activity. Only the amount of variance associated with Ascendance and Sociability remained the same for LB and for LA. We can offer no reasonable explanation at this time for this interesting but perplexing event.

c. *Discussion participation or attempted discussion leadership activity* (PN) was greater, the greater the Verbality and Ascendance and Sociability of the participants. But it had less in common with Sorority Leadership Potential, as such, than successful leadership activity (LA). In other words, there appeared to exist an X quantity which when present among socially bold and verbal participants enabled them to be successful discussion leaders and to some extent to have sorority leadership potential, but when absent enabled them only to attempt with much less success to gain leadership status. To search for this X quantity, an index of the aborted or unsuccessful leadership behavior exhibited by participants of the first discussion was calculated by subtracting each subject's standard score of successful leadership activity (LA) from her standard score of attempted leadership activity (PN). Inspection of the correlations between aborted leadership status (PN-LA) and the other assessment variables suggested that the X quantity might have been objectivity or freedom from hypersensitivity (ZO), which correlated $-.44$ with aborted leadership activity. It might also have been cooperativeness or freedom from criticalness or intolerance (ZP), which correlated −.32 with aborted leadership activity.

d. The results concerning discussion efficiency are fairly clear yet somewhat comic in their implications and relations with stereotypes about women as conversationalists. A correlation of −.56 was found between the femininity of participants and the efficiency of their discussion! Apparently the way to obtain an effective discussion among women is to select women with masculine temperament who have had some university-wide extracurricular experience and activity.

e. Probably the most surprising result of the correlational and factorial analysis was the tendency of the stimulated recall variables to be fairly independent of other measures and to a great extent of each other. In general, most of the reliable variance in thought processes was independent of behavior and quite unique to the specific thought processes involved. However, what communality was observed appeared quite meaningful and useful. Two and possibly three thought-process factors emerged to account for most of the common variance found among the stimulated recall variables. These were: Critical, Relevant Thinking; Group-Process Thinking; and possibly Withdrawal-from-the-Field Thinking.

VI. SUMMARY

Data on 41 variables were collected for 140 sorority women representing seven sororities. The five types of data included: (*a*) biographical information, (*b*) peer ratings (of own sorority sisters), (*c*) self-ratings, (*d*) paper-and-pencil test scores, and (*e*) situational test scores.

Primary interest was focused on increasing our understanding of the various situational test procedures employed and situational and personality factors associated with leadership among sorority girls.

Leadership was defined as a phenomenon in which one group member's behavior was directed toward the goal of changing the behavior of another member or members. The leadership potential of the girls was assessed for three different situations: in nonsorority university activities, in sorority activities, and in initially leaderless discussions. On-campus leadership potential was estimated from campus leadership status or campus positions of worth held by the subjects. Sorority leadership potential was determined on the basis of esteem ratings of sorority sisters and from leadership positions held. Discussion leadership potential was gauged in terms of the extent to which each subject was judged by two observers as having exhibited successful leadership activity.

The matrix of intercorrelations among the 41 variables was subjected to a multiple centroid factor analysis using a group clustering method as a computational short cut.

As a result of these correlational and factorial analyses, the propositions listed below concerning leadership and associated concepts of importance such as participation, esteem, and status were substantiated to some degree. Because of the exploratory nature of this study, these conclusions, unless supported by other inductive or deductive evidence, should be considered tentative or suggestive. In evaluating the evidence, it may be helpful to keep in mind that a correlation of .21 is significant at the 1 per cent level and a correlation of .16 is significant at the 5 per cent level of confidence for 140 cases.

A. BIOGRAPHICAL INFORMATION

1. Older sorority women with more university experience are more active in extracurricular pursuits ($r = .60, .53$), are higher in university and sorority leadership status ($r = .43, .41, .41, .33$), tend to be more esteemed by their sorority sisters ($r = .20$ to $.28$), and tend to be less motivated to be a leader in their sororities ($r = -.23, -.17$).

2. Both university and sorority leaders tend to participate in more extracurricular activities than nonleaders ($r = .62, .36$).

3. Women with a previous history of extracurricular interest and activity tend to be more active and effective in initially leaderless group discussions ($r = .32$).

4. Personality differences tend to be more discriminating among university leaders and nonleaders than among sorority leaders and nonleaders. Women higher in university leadership status tend to be more socially bold ($r = .27$), emotionally stable ($r = .21$), and active and effective in leaderless discussions ($r = .36$).

5. Sorority leadership status tends to be slightly less associated with various measures of ability than does university leadership status.

6. A low positive correlation ($r = .23$) exists between university and sorority leadership status.

7. Sorority leaders must conform to some extent with the cultural norm that women be socially submissive. The same requirement does not apply to women campus leaders.

B. PEER RATINGS (OF OWN SORORITY SISTERS)

1. Peer ratings of sorority esteem provide a pure measure of leadership potential. Almost all their reliable variance is associated with Sorority Leadership Potential.

2. Esteem is slightly more associated with personal characteristics than is sorority status, although esteem and status are highly correlated.

3. Highly visible sorority members are of high Sorority Leadership Potential ($r = .54$) and tend to be slightly Ascendant and Sociable ($r = .19$).

4. Visibility in the sorority does not correlate, as expected, more highly with status than it does with esteem.

5. Sorority women who are overtly socially adjusted tend to be high in Sorority Leadership Potential ($r = .54$), objective ($r = .41$), friendly ($r = -.39$), less socially bold ($r = -.18$), and less socially interested ($r = -.24$).

6. A socially bold sorority member who lacks esteem tends to be highly maladjusted socially.

C. SELF-RATINGS

1. Self-esteem appears to be a compromise between beliefs about actual leadership potential and desired leadership potential. Leaders, masculine efficient women, socially bold women, women more motivated to attain leadership status, and more verbal women are higher in self-esteem.

2. Sorority members tend to be more motivated to attain university leadership status than sorority leadership status.

D. SITUATIONAL TEST SCORES

1. Variations among individuals in success as a leader in an initially leaderless discussion are a function of Verbality

($r = .51$), Ascendance and Sociability ($r = .39$), Sorority Leadership Potential ($r = .24$), Intellectualism ($r = .21$), and University Experience and Activity ($r = .18$).

2. Attempted leadership activity in a leaderless discussion is a function of Verbality ($r = .45$) and Ascendance and Sociability ($r = .39$).

3. Aborted or unsuccessful leadership activity in a leaderless discussion is greater among hypersensitive ($r = -.44$) and critical, intolerant sorority women ($r = -.32$).

4. The efficiency of initially leaderless group discussions tends to be lower, the more feminine in temperament the women participants ($r = -.56$) and the less university experience they have had ($r = .28$).

REFERENCES

1. Bass, B. M. An analysis of the leaderless group discussion. *J. appl. Psychol.*, 1949, **33**, 527-533.
2. Bass, B. M. Situational tests: I. Individual interviews compared with leaderless group discussions. *Educ. psychol. Measmt*, 1951, **11**, 67-55.
3. Bass, B. M. Differential response patterns in initially leaderless discussions related to discussion and external status. *Amer. Psychologist*, 1951, **6**, 511. (Abstract)
4. Bass, B. M. A psychological theory of leadership. Baton Rouge: Louisiana State University, 1953. (Mimeographed)
5. Bass, B. M., McGehee, C. R., Hawkins, W. C., Young, P. C., & Gebel, A. Personality variables related to leaderless group discussion behavior. *J. abnorm. soc. Psychol.*, 1953, **48**, 120-129.
6. Bass, B. M., & Wurster, C. R. Effects of company rank on LGD performance of oil refinery supervisors. *J. appl. Psychol.*, 1953, **37**, 100-104.
7. Bloom, B. The study of conscious thought processes by the method of stimulated recall. Mimeographed manuscript, Univer. of Chicago.
8. Cattell, R. B. *Description and measurement of personality*. Yonkers-on-Hudson: World Book Co., 1946.
9. Cattell, R. B. *Factor analysis*. New York: Harper, 1951.
10. Fleischman, E. Leadership climate and supervisory behavior. Unpublished doctoral dissertation, Ohio State Univer., 1951.
11. Guilford, J. P., & Zimmerman, W. *Manual of the Guilford-Zimmerman Temperament Survey*. Beverly Hills, Calif.: Sheridan Supply Co., 1949.
12. Hemphill, J. Leadership in small groups. Mimeographed manuscript, Personnel Research Board, Ohio State Univer.
13. Klubeck, S., & Bass, B. M. Differential effects of training on persons of different leadership status. *Hum. Relat.*, in press.
14. McClelland, D. C. Personality. New York: William Sloane, 1951.
15. Peterman, J. N. Verbal participation, its relation to decision satisfaction and leader function in decision-making groups. Paper read at Midwest. Psychol. Ass., Chicago, April, 1949.
16. Smith, M. Control interaction. *J. soc. Psychol.*, 1948, **28**, 263-273.
17. Stogdill, R. M. Personal factors associated with leadership. A survey of the literature. *J. Psychol.*, 1948, **25**, 35-71.
18. Super, D. *Appraising vocational fitness.* New York: Harper, 1949.
19. Wherry, R. H. & Fryer, D. H. Buddy ratings: popularity contest or leadership criteria? *Personnel Psychol.*, 1949, **2**, 147-160.
20. Williamson, E. G. The group origins of student leaders. *Educ. psychol. Measmt*, 1949, **11**, 603-612.
21. Williamson, E. G., & Hoyt, D. Measured personality characteristics of student leaders. *Educ. psychol. Measmt*, 1952, **12**, 65-78.
22. Wurster, C. R., & Bass, B. M. Situational tests: IV. Validity of leaderless group discussions among strangers. *Educ. psychol. Measmt*, 1953, **13**, 122-132.

Vol. 67, No. 17 Whole No. 367, 1953

Psychological Monographs: General and Applied

The Development of Phobias in Married Women[1]

A. Stanley Webster

University of Tennessee and Eastern State Hospital

I. INTRODUCTION

In considering the etiology, status, and consequents of abnormal fear reactions there is not only a dearth of experimental data, but also lack of clarity in use of the terms involved. Generally, abnormal fear reactions are briefly mentioned in passing as being self-evident. Few hypotheses regarding causal factors and their methods of combination have been attempted. Few of the results of studies concerning phobias seem to be objectively verifiable. Lack of clear, certain knowledge of the phobias seems particularly regrettable, since phobias are generally considered to be a link between the neuroses and the psychoses.

It is hoped that the results of this study may cast some light upon the etiological factors involved in the development of the phobic syndrome.

II. METHODS AND PROCEDURE

A. General Approach to the Problem

In beginning a discussion of the problem, of first importance is the matter of determining a practicable method for dealing with phobic patients. In this instance, the case study method was chosen. Our purpose is to seek by careful inspection those factors which seem etiologically significant in the patient's life history, or at least to see if any threads of commonality recur throughout the histories of a number of patients. Even if several patterns of events recur throughout a number of histories of patients, however, this does not guarantee that we will be able to attach the proper significance to them, or be able to tell what etiological part they play in an illness. But the point is that we do not yet know what events might occur or recur, and it is for these things we must first seek.

It is the purpose of this study, therefore, to determine whether, to what degree, and in what order certain variables appear in a sample of psychopathological patients—these variables being ones which have been mentioned by one author or another as being etiologically significant or have been chanced upon by the present writer.

B. Research Population

The population used in this experiment consists of three groups of patients who have received at least three months' psychiatric treatment. Each group is composed of 25 adult white females. One group of 25 is composed of phobic patients, i.e., women whose presenting complaint upon admission to treatment was an abnormal fear reaction. It should be mentioned that all phobic patients were agoraphobics. They were afraid to go out on the streets alone and many were afraid to remain at home alone. Do all phobics eventually become agoraphobics, or are all phobics who come to psychiatrists for help agoraphobics, or is this a selected sample? This question cannot be answered at this time any more than the obvious question which arises—are only females phobic, or are all female phobic patients married? This study is not in a position to make a valid reply to either question, but the literature which was consulted reveals few cases of unmarried female phobics or of male phobics. The second group is composed of patients who, after the prescribed three months of treatment or longer, were given a diagnosis of conversion hysteria; and the third is composed of patients whose diagnosis, after three or more months, was anxiety neurosis. The "Summary of Research Population" presents the patients used in this study.

---

[1] Submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy, in the faculty of Philosophy and Psychology, University of Tennessee, 1952.

SUMMARY OF RESEARCH POPULATION

Diagnosis	*N*	Mean Age	Marital Status
Phobia	25	30.92	100% married
Anxiety Neurosis	25	32.48	92% married 4% widowed 4% separated
Conversion Hysteria	25	31.80	88% married 4% widowed 4% separated 4% divorced

The case histories of these 75 patients were secured from psychiatrists in private practice, in federal agencies, and in city and state institutions. The case histories included psychiatric, psychological, and physical examinations in addition to neurological evaluations and, in certain instances, laboratory work or operative procedures. In addition to the intake interviews—some of which were made by psychiatrists, some by social workers, and still others by psychologists—there were notes made by the therapist during the course of treatment. Obviously, no two histories contained the same factual material, nor was the material in any particular order. In other words, the "case history" of each patient consisted of her "file" just as it was lifted out of a filing drawer, with no rearrangements, summarization, or shifting of material therein. Physicians were simply asked if they would release a "file *in toto*" on any patients they might have whose diagnosis was "phobia," "anxiety neurosis," or "conversion hysteria," without disturbing the material in any way. They were told that the material would be kept highly confidential except for the fact that a person unacquainted with any of the patients would read the file to make certain judgments, and that the person doing this would probably be a physician. It was further indicated that as soon as the judgments had been made, the file would be returned intact. Sample abstracts of the case histories of three phobic patients may be seen in Appendix A.[2]

C. DESIGN AND PROCEDURE

The case histories previously mentioned were gathered for the purpose of having judgments made concerning them. Therefore, a rating scale was devised (see Appendix B) which would enable a rater to estimate the degree of presence or absence (Quite Marked, Moderate, Barely Noticeable, Not Present) and the order of appearance of the following points concerning the patients:

---

[2] Complete case abstracts and ratings are available in Webster, A. S., The development of phobias in women. Unpublished doctoral dissertation, Univer. of Tennessee Library, 1952.

1. A father figure and his behavioral characteristics, i.e., the physical or emotional presence or absence of a male figure whose personality picture is adequate or inadequate, and its type.

2. A pattern of dominant overprotection by the mother; i.e., the bestowal of much affection on the child when good, and the withdrawal or threatened withdrawal of it when the child is bad.

3. Marriage to an older man; i.e., the number of years by which the husband's age exceeds the patient's.

4. The husband's behavioral characteristics; i.e., the adequacy or inadequacy of his personality and its type.

5. Castration and/or castration fear; i.e., the symbolic fear or surgical castration procedures.

6. Pregnancy and the rejecting of it; i.e., actual childbirth or its antecedent manifestations which are on an organic basis, and the actual or symbolic rejection of such a condition.

7. Frigidity; i.e., the amount of warmth shown by the patient toward her husband.

8. The husband's withdrawal; i.e., departure, threatened departure, or emotional withdrawal.

In order to ensure reliability, two judges were employed to make the judgments and one of the judges, after an interval of several months, repeated his judgments. The first judge, an obstetrician-gynecologist with psychiatric training, was informed of nothing concerning the investigation other than the fact that he would have to make judgments from case history material sent to him, that he had to make up his own subjective norms for each of the points on the rating sheet and use the same norms in judging each case, and that he should then write out as clearly as he could the norms which he had used in making his evaluations. Since the material sent to him contained *all* of the available material concerning the patient, obviously the diagnostic labels were also contained in the material. For this reason, so that no bias could develop if it chanced that he received several files on the same type of patient, the 75 cases were arranged according to a table of random numbers and sent to him in this order. The norms which this first judge used were then given to the second judge a year later in order that he might use approximately the same judgmental norms. The second judge, however, did not have the entire case material on which to make his judgments, but only the abstracts made by the writer, three of which appear in Appendix A. (Obviously, it was impossible to keep the case histories a year; after they had been returned by the first judge, an abstract was made of each case and the original material returned.) Considering this point in the procedure, if, when both intrarater reliability (reliability between Judge 1 and himself after making judgments of the same material several months apart) and interrater reliability (reliability between Judge 1 and Judge 2) were computed, the reliability coefficients approached each other numerically, the use of abstracts rather than the full case material would be indicated, with a marked economy of time for the person reading the material in order to make the judgments.

Judge 2 was a third-year graduate student in clinical psychology. The norms were supplied to him, with particular emphasis upon the fact that he was to use the same norms as those arrived at by Judge 1. These norms were as follows and will be referred to hereafter as "variables," corresponding to the numbers given below:

1. *Father Figure*

Quite Marked—The patient's father must have been a salesman, buyer, or had some similar employment which required him to be absent from the home an average of four nights a week; or should be an alcoholic, gambler, faithless to his wife, and rejecting of the children.

Moderate—The patient's father may have been employed in any of the activities mentioned above but been absent from home only 2-3 nights a week and should seem from the case history to be "fairly adequate" as a father and husband.

2. *Dominant Overprotection*

Quite Marked—The patient's mother should be most solicitous of the child's welfare, rewarding her often without good reason, and rejecting or threatening to reject her or actually telling her she would not love her any more if she did not "behave."

Moderate—The patient's mother should do the above-mentioned activities, but to such a degree that it is mentioned by only one or two persons in the statements compiled in the folder.

3. *Marriage to an Older Man*

Quite Marked—The patient should have married a man seven or more years older than herself.

Moderate—The patient should have married a man 4-6 years older than herself.

4. *Castration and/or Castration Fear*

Quite Marked—The patient should have had a panhysterectomy, subtotal hysterectomy, partial hysterectomy, or a bilateral salpingectomy.

Moderate—The patient should have had a unilateral salpingectomy or many dreams of pursuit in which she was the pursued party, while the pursuing party carried some sharp object and meant her harm—guns are not acceptable, nor are sticks, rocks, or any form of club. The patient may also be obsessed with the idea of not going to sleep unless all of her limbs are covered with a sheet or blanket or unless all objects on her dresser are arranged so that she feels there is no danger of anything falling off.

5. *Pregnancy and its Rejection*

Quite Marked—The patient's actual statement that she did not wish to become pregnant; attempts at abortion, suicide, or self-maiming.

Moderate—Deep depression on the part of the patient during pregnancy, and statements of persons other than the patient concerning her dislike for being pregnant.

6. *Frigidity*

Quite Marked—The patient's actual statement that she is frigid and dislikes her husband.

Moderate—The patient's husband states that she is frigid.

7. *Husband's Withdrawal*

Quite Marked—Actual separation, initiated by the husband; or continual threatening of separation on his part, or extreme emotional aloofness from the patient.

Moderate—Occasional threatening of the husband to separate from, divorce, or otherwise leave the patient; or poor emotional relationship between the patient and her husband.

8. & 9. *Other Characteristics of Patient's Father and Husband*

The other characteristics of the patient's father (8) and husband (9) are in terms of diagnostic categories and the appropriate classification is made depending upon the history of either given. Thus a description of a shy, withdrawn individual would be "schizoid," while a faithless, alcoholic gambler would be termed "psychopathic," etc.

These norms will be further discussed in a section concerning the limitations of this design.

Thus, judgments were made on the same set of randomly presented material by two judges using approximately the same norms, although it must be remembered that the second judge used only abstracts instead of the full case material.

This study was not amenable to elaborate statistical analysis, primarily because there is no known instrument which would accurately measure that which this study considers. For this reason, statistical techniques are purely for descriptive and summarization purposes. For such purposes, the methods of percentages and chi square were employed to determine whether there were significant differences between the phobic group and each of the other two groups in terms of number of symptoms present and the intensity of such symptoms.

D. LIMITATIONS OF THE DESIGN

It is obvious that the norms used in this study are subjective. The first judge was not even able to list those events in the case history which would lead him to a judgment of "Barely Noticeable." It might be argued that, even though he had difficulty in expressing those factors which led him to his judgments, if he and the second judge arrived at the same judgments, then however poor the norms may be, they actually enabled two people to arrive at the same conclusions. This is not a valid point, however, since many people, if trained in similar fashion, might arrive at the same conclusions with poor norms, while were the norms more adequate, they might have arrived at different conclusions. In any case, this inadequacy with respect to norms should be recognized; it simply points out that a weakness in quantification is inherent in the nature of this study.

It is also to be noted that the groups are unmatched in variables which may be significant, although it is assumed that randomization minimizes this weakness.

Finally, it should be mentioned that the design is not practical—judging by the fact that the first judge, utilizing case history material *in toto,* took over 160 hours. This weakness could be offset if it were found that abstracts could serve the same purpose with as great reliability as the entire case record.

III. RESULTS

A. RELIABILITY OF RATINGS

Reliability coefficients were computed between the first and second ratings of Judge 1, and the first ratings of Judge 1 and those of Judge 2. These were product-moment correlations with Sheppard's correction. Contingency coefficients were also computed for the comparison of judgments between the diagnoses for the "other characteristics of father and husband." For statistical purposes, any diagnosis other than "normal" was considered "unstable." Since there is only a twofold division, some statistic is needed to compare stable vs. unstable, and thus a contingency coefficient was chosen. The three groups of patients will be hereafter referred to as P, CH, and AN for the phobic, conversion hysteric, and anxiety neurotic, respectively. Table 1 presents the coefficients computed. It is to be noted that the maximum contingency coefficient possible with a fourfold table is .866. In the case of "marriage to an older man," the ratings were identical for all three judgments. Thus no reliability coefficient could be computed.

Since the interrater correlations are quite high, we need employ only one judge's set of ratings in computing the significance of differences between the three groups with respect to variables 1-9, described previously. Therefore, only the first ratings of Judge 1 will be used in the computations to follow.

B. RATINGS OBTAINED OF INDIVIDUAL VARIABLES

Tables 2–8 are divided into two parts, one of which shows the percentage of occurrence of each variable for the groups. The other part shows the chi squares and level of significance at which we can assume that these are not chance

TABLE 1

OVER-ALL RELIABILITY COEFFICIENTS

Variable	Intrarater	Interrater	p
	Correlation Coefficients		
1. Father figure	.92	.91	.01
2. Dominant overprotection	.92	.91	.01
3. Marriage to an older man	—	—	—
4. Castration	.97	.93	.01
5. Rejection of pregnancy	.88	.86	.01
6. Frigidity	.88	.87	.01
7. Husband's withdrawal	.88	.86	.01
	Contingency Coefficients		
8. Husband's characteristics	.83	.80	.01
9. Father's characteristics	.82	.80	.01

TABLE 2

INCIDENCE OF LACK OF AN ADEQUATE FATHER FIGURE

Group	% Incidence	Incidence/Total	Comparison	χ^2	df	p
P	100	25/25				
AN	52	13/25	P-AN	12.042	1	.001
CH	56	14/25	P-CH	10.386	1	.001

results. The chi squares were computed with Yates's correction from combining the categories "Quite Marked" and "Moderate," as opposed to the combined categories "Barely Noticeable" and "Not Present." Even though it can be seen that in the latter categories there was some "presence" of the variable, to the extent that it was noticeable, for the sake of statistical expediency the two gross categorizations were so set up. The same methods of combination and computation hold true for the remainder of the tables in this section, so it can be seen that whenever a column is labeled "% Present," it means present to a "Quite Marked" or "Moderate" degree.

1. *Lack of an adequate father figure.* The indications are that patients in the P group show a significantly greater tendency toward lacking an adequate father figure than does either of the other two groups.

2. *Dominant overprotection.* The indications are that patients in the P group show a significantly greater tendency

TABLE 3

PRESENCE OF DOMINANT OVERPROTECTION

Group	% Present	Presence/Total	Comparison	χ^2	df	p
P	96	24/25				
AN	44	11/25	P-AN	13.714	1	.001
CH	44	11/25	P-CH	13.714	1	.001

TABLE 4

PRESENCE OF MARRIAGE TO AN OLDER MAN

Group	% Present	Presence/ Total	Comparison	χ^2	*df*	*p*
P	56	14/25	P-AN	1.288	1	Not significant
AN	36	9/25	P-CH	1.288	1	Not significant
CH	36	9/25				

TABLE 5

PRESENCE OF CASTRATION OR CASTRATION FEAR

Group	% Present	Presence/ Total	Comparison	χ^2	*df*	*p*
P	92	23/25	P-AN	25.961	1	.001
AN	12	3/25	P-CH	16.638	1	.001
CH	32	8/25				

toward having experienced dominant overprotection on the part of the mother than does either of the other two groups.

3. *Marriage to an older man.* The indications are that marriage to an older man occurs no more often in any one of the three groups than it does in either of the other two.

4. *Castration or castration fear.* The indications are that patients in the P group show a significantly greater tendency toward having experienced castration or castration fear than does either of the other two groups.

5. *Rejection of pregnancy.* The indications are that patients in the P group show a significantly greater tendency toward having rejected pregnancy than do patients in either of the other two groups.

6. *Frigidity.* The indications in the case of this variable are that the P group shows a significantly greater tendency toward being frigid than does the CH group, but it is questionable whether an adequate differentiation may be made between them and the AN group.

7. *Husband's withdrawal.* The indications are that the husbands of patients in the P group show a significantly greater tendency toward withdrawal from the patients than do husbands of patients in either of the other two groups.

8. *Other characteristics of the husband.* In this table, as in Table 1, the

TABLE 6

PRESENCE OF REJECTION OF PREGNANCY

Group	% Present	Presence/ Total	Comparison	χ^2	*df*	*p*
P	92	23/25	P-AN	13.220	1	.001
AN	36	9/25	P-CH	25.961	1	.001
CH	12	3/25				

TABLE 7

PRESENCE OF FRIGIDITY

Group	% Present	Presence/ Total	Comparison	χ^2	df	p
P	92	23/25				
AN	64	16/25	P-AN	1.780	1	.20
CH	44	11/25	P-CH	9.191	1	.01

terms *normal* and *unstable* are applied to the judgments made. Table 9 additionally presents a breakdown into nosological categories of the percentages of occurrence in the opinion of the rater.

The indications are that the husbands of patients in the P group show a significantly greater tendency to be unstable or inadequate than do the husbands of patients in either of the other two groups.

TABLE 8

PRESENCE OF HUSBAND'S WITHDRAWAL

Group	% Present	Presence/ Total	Comparison	χ^2	df	p
P	100	25/25				
AN	24	6/25	P-AN	26.775	1	.001
CH	60	15/25	P-CH	8.82	1	.01

9. *Other Characteristics of the Father.* The presentation of this table is identical with that of Table 9.

The indications are that the fathers of the patients in the P group show a significantly greater tendency to be unstable or inadequate than do the fathers of patients in either of the other two groups. It may seem at first glance that these tables (Tables 9 and 10) present incredible percentages, but it must be mentioned again that *descriptive* terms only were used.

It requires mention that in five instances in the P group, three instances in the CH, and six instances in the AN group, there was insufficient information from which to make a judgment. Therefore, in the computations of the percentages and chi squares, these omissions were duly noted and the N for the particular variable under consideration was reduced by the appropriate number. In

TABLE 9

PRESENCE OF AN UNSTABLE HUSBAND

Group	% Present	Groups	χ^2	df	p
P*	96	P-AN-CH	43.315	4	.001
AN**	8	P-AN	35.336	1	.001
CH†	8	P-CH	35.336	1	.001

* 92% of these men were adjudged psychopathic, 4% neurotic, and 4% normal.
** 88% of these men were adjudged normal and 12% psychopathic.
† 88% of these men were adjudged normal and 12% psychopathic.

TABLE 10

Presence of an Unstable Father

Group	% Present	Groups	χ^2	df	p
P*	96	P-AN-CH	39.511	4	.001
AN**	16	P-AN	35.562	1	.001
CH†	20	P-CH	26.600	1	.001

* 87.5% of these men were adjudged psychopathic, 8.3% schizophrenic, and 4.1% normal.
** 81.8% of these men were adjudged normal, 13.6% psychopathic, and 4.5% schizophrenic.
† 75% of these men were adjudged normal, 20% psychopathic, and 5% schizophrenic.

no patient's case, however, were more than two judgments omitted by the rater, and these omissions occurred on only on one occasion; the remainder of the single omissions were widely scattered.

C. Analysis of Sequences of Variables

Table 11 was constructed to show the percentage of times that one varible appears in the record prior to others. This table is incomplete since some variables cannot make their appearance prior to the occurrence of others (rejection of pregnancy cannot appear prior to marriage) and also because only those percentages which, by inspection, appeared large, were computed. The remaining two groups (CH and AN), excluding those variables which cannot occur prior to the occurrence of another, show percentages ranging only in the twenties, with the exception of lack of an adequate father figure, which precedes dominant overprotection 92 per cent of the time in the case of the AN group, and 84 per cent of the time in the case of the CH group. In only three instances could judgments as to the order of precedence not be made in the P group (because of incompleteness of the record), while there are 78 such instances in the AN group and 86 in group CH.

TABLE 11

Analysis of Sequences for Group AH

Lack of an adequate father figure precedes dominant overprotection	100% of the time
Rejection of pregnancy precedes castration or castration fear	65% of the time
Frigidity precedes rejection of pregnancy	91% of the time
Frigidity precedes husband's withdrawal	70% of the time
Frigidity precedes castration or castration fear	48% of the time

Table 11 indicates some fairly well-regularized relations in the P group, most notable of which is that lack of an adequate father figure occurs before dominant overprotection on all occasions. This occurs frequently, however, in both of the other groups. The remaining four percentages shown indicate a trend toward being present in the P group. No such trend is apparent in either of the other two groups.

It should be noted that all patients in the P group were judged as possessing to some degree all of the traits of the variables under consideration. Those instances which lead to percentages lower than 100 are possibly due to insufficient information for making a judgment. If those instances in which there was an insufficient amount of information given to make a judgment were discarded and the "Barely Noticeable" category were included in the percentage computations, all the patients would have shown presence of all the variables, with the exception of marriage to an older man. Statistical expediency, however, dictated that the "Barely Noticeable" category should not be included.

A final addendum should be made concerning the results found. This concerns one additional statistical procedure —a chi square representing a finding which did not appear, and which was not suspected, until the results of the judgments were in. On the judgment sheet, as shown in Appendix B, there is a space for the judge's observation as to whether any other variables occur which he feels to be of importance in the case history. A number of symptoms were observed by the judge and placed in these spaces, but the one which occurs with great frequency is "exposure to therapist"—that is, the patient in some way exposes her breasts, legs, or genitalia to the person treating her. This was noted in 18 of the 25 phobic cases, in one case of anxiety neurosis, and in two cases of conversion hysteria. Among the three groups for exposure to therapist, the chi square was 93.381. This is significant at the .001 level, indicating it appears likely at a very high degree of confidence that the finding is a real one. Thus, the fact that exposure to the therapist occurs a significantly greater proportion of the time in the P group as compared to the other two groups does not seem to be an artifact. It should be noted, however, that the judge may not have noted with equal emphasis the presence or absence of this symptom in all cases, although it must be remembered that this chi square represents only the first judgments of Judge 1, and also that the cases were presented to him in random order. Nevertheless, it must be regarded more tenuously than the other chi-square findings. Expressed in terms of percentages: (*a*) 24 per cent of the P group "exposed" themselves to the therapist as compared with 1.3 per cent of the CH group; (*b*) 9.3 per cent of the P group did not "expose" themselves to the therapist as opposed to 32 per cent of the AN group and 30.6 per cent of the CH group.

IV. DISCUSSION

The results shown in the preceding section should be considered as tentative and preliminary to further research. Therefore, it should be borne in mind that to say that the findings have differentiated phobias (anxiety hysteria) from anxiety neuroses and conversion hysteria, or that it is concluded that we have differentiated anxiety hysteria from anxiety neuroses and conversion hysteria is to say little, since this is accomplished in the presenting complaints. One of the prime reasons for the dearth of research in the phobic field is that the diagnosis is immediately established. It seems, however, that something more than differentiating by virtue of diagnosis has been accomplished. In these cases, a differentiation has been made on the basis of the occurrence of certain variables, so that, even in the absence of knowledge of presenting complaints or of the diagnosis, the existence of a phobia could be deduced. Such a condition exists at least in the cases presented in this study.

The results indicate quite clearly that lack of an adequate father figure, dominant overprotection by the mother, inadequacy of the patient's husband, castration or castration fear, the rejection of pregnancy, frigidity, and the husband's withdrawal regularly occur in these phobic women. Does this mean that they are causal factors? If they were not, why did they not appear with equal frequency or strength among the patients who did not have phobias? There is every statistical and clinical indication that such occurrence is not on the basis of chance. It cannot be argued, for example, that the variables might be the result of the phobia rather than a cause, since they do not occur temporally after the development.

Clinically, the phobic patient presents too many of the combined symptoms of both anxiety neurosis and conversion hysteria to justfy the view that these three syndromes come from totally different universes. Yet they appear quite different in this study. This is due to the fact that particular variables are causal factors in the development of the phobia—which cannot develop without them—while they simply contribute to the general situation in the other two groups. Thus, the clinical appearance of combined symptoms in the phobic, from both anxiety neurosis and conversion hysteria, can be reconciled with the statistical appearance of disparate variables in the phobic group as compared with each of the other two groups.

The variables used in this study have been referred to as "variables" rather than "events" in a life history because they have appeared with such regularity in phobic cases other than those used in this study, while the term "event" implies a happening which may or may not be of importance. Thus, "variable" is used advisedly to indicate a happening of some particular importance. The variables used in the present study are also found in almost all other case histories relating to female phobias, but they have been almost universally ignored. The chief reason for this neglect seems to be that the facts are isolated, apparently have little superficial relation to each other, and are either buried in seemingly more important facts, or in esoteric terminology. When these variables do not appear, it is usually a result of insufficient information, poor history-taking methods, or bias on the part of the examiner. It might be argued that it is easy to explain away anything by saying that there is insufficient information given to determine whether or not a symptom is present. This is a valid objection. The facts in this case, however, do not support such a contention, since in all cases if the possibility of existing were present, the variable itself was present—that is to say, in no cases where there was sufficient information given to make a judgment was a variable absent. Allowing for the "insufficient information" factor, *all* variables were present in *all* cases, disregarding "marriage to an older man."

If they are causal, what relation can these variables have to the development of phobic reactions in women? The hypothesized relationship is that in the absence of a father figure, the patient develops ambivalent feelings toward the mother because of her dominant overprotection, which generally causes anxiety and feelings of insecurity—that is, the patient is unable to develop her potentialities, to experience the capacity for self-realization. Yet the anxiety and ambivalent feelings toward the mother figure push the patient to leave the mother. She does this in socially acceptable fashion, however, by marrying. It is felt that she chooses an inadequate person both because of desire to displease the mother and inability to choose wisely because of the restrictions which have been placed on her by the mother; i.e., her dependency on the mother precludes opportunity for wide experience. The patient is frigid because of immaturity and because of fear of pregnancy for which she is totally unequipped. She herself wishes to be dependent, so it can hardly be expected that she can permit someone else to be dependent upon her. Nor is her husband adequate for her dependency needs, which is one of his reasons for leaving the patient. That is, the husband, originally a "weakling," who himself needs someone upon whom he can be dependent, finds out that not only is his wife not such a person, but also that she has robbed him of his manhood by not wishing to have a child.

Castration is here conceived of not necessarily as a hysterectomy, but as any act which deeply traumatizes the patient. The persons who are insecure are fearful of being hurt. It is for this reason that they develop the phobia, literally in order to protect themselves. They wish to be dependent and perhaps also to keep the husband near them to satisfy their dependency needs. These acts, however, if they have not already alienated the husband, will do so soon—being unable to cope with a person more sick than himself, he will leave.

Many authors have conceived of castration as playing an important part in the development of anxiety hysteria or phobias. Yet, from the results of this study it does not seem more important than the other symptoms cited; it seems to be simply one link in a chain of events, all links of which must be present before the chain is complete. It is felt, however, that castration may be the genotype for a variety of events; that it may have other forms than surgical castration; and that if all other symptoms deemed necessary to the development of the phobia save castration are present, any severe trauma to the patient's self-conceptualization may cause a phobia to develop. This point will be discussed further.

The list of variables compiled throws light on the reason for the great difficulty in extinguishing the phobia. None of the variables except those pertaining to the husband is reversible. That is, lack of an adequate father figure, dominant overprotection, and castration are irreversible phenomena. The husband's leaving, his behavioral characteristics, and the patient's frigidity and rejection of pregnancy (the last only if castration is not complete) can conceivably be reversed or changed and thus break the chain of symptoms leading to the phobia.

Thus, *the husband's position can hardly be overestimated since he provides the only link in the chain which can be attacked.* Clinically, this is fairly easily accomplished since the husband himself is immature, has suffered a blow to his ego, and would like to have someone on whom he can be dependent, even more now than prior to his wife's development of the phobia. The person upon whom he can feel dependent is quite conceivably a therapist. In most instances, however, the husband is hostile toward his wife's therapist for "stealing" her from him, and so impugns the therapist vehemently. If this is not enough, he must admit his failure to manipulate his own life. If he can be gotten to come for interviews to a therapist other than his wife's therapist, there is much less conflict and his dependency needs can more easily be satisfied in the presence of one who knows nothing of him than with one who is familiar with his circumstances. It is interesting to note that in this study only those patients whose husbands came for treatment showed any degree of improvement at the time their case history was secured. The point cannot be overemphasized that the husband is the therapist's only means of entrance into a closed and self-sustaining system.

The final point to be made involves a question which one asks oneself after seeing the results of this study—do only those female patients who present the variables enumerated develop phobias?

The answer to the question is a difficult one in view of the fact that (apart from those cases where no judgment could be made) all patients presented all variables cited except marriage to an older man. It seems likely, however, that patients can develop phobias as a result of one tremendously severe trauma or several sequentially similar traumas. Such phobias, however, dissipate independently over a fairly short period of time or quite quickly with psychotherapy, in contrast to the phobias cited in this study, and generally are based upon a quite realistic fear or generalization of some realistic fear. Moreover, at this point we can see that phobias may not really be irrational fears at all, nor intense fear-produced responses of childhood, nor illogical fears. If full account is taken of the various factors present in each case of phobia, there is every reason to hope that, with further research, a definite causal sequence can be established to remove the last doubt that these manifestations are not irrational or illogical fears.

V. SUMMARY AND CONCLUSIONS

This study was undertaken with the hope that the results might cast some light on the etiological factors involved in the development of phobias in married women.

Three groups of white female patients were selected. Each group of 25 patients was comprised of married women with the same diagnostic label. One group was composed of phobics, one of anxiety neurotics, and one of conversion hysterics. These patients had received a minimum of three month's psychotherapy. The case history of each of the patients was secured. Each case history contained medical, psychiatric, psychological, and, in some cases, neurological reports and operative procedures. A rating sheet was devised in order to permit a judge to rate the order of appearance of certain variables, together with their intensity. Those points to be judged were presence or absence of: an adequate father figure, dominant overprotection by the mother, marriage to an older man, castration of the patient, frigidity of the patient, rejection of pregnancy by the patient, departure or threatened departure on the part of the husband, adequacy of the husband. Rating sheets and randomly ordered case histories were presented to a judge for the purpose of making such ratings and enumerating any other factors he considered of importance. Several months later he was again required to make the same judgments. One year later abstracts of the cases were presented to a second judge together with rating sheets in order that he might also make judgments. Statistical analysis of the data by means of chi square and percentages was undertaken to determine the significance of the relationship of the variables to the three groups of patients. Further statistical analysis was undertaken to establish the reliability of the three sets of judgments. For this purpose Pearson product-moment correlations and contingency coefficients were employed. A presentation of the findings was followed by a discussion of the results and implications of this study, together with a proposed explanation of the importance of the variables for development of phobias.

From this study it is concluded that

the variables under consideration—namely, (*a*) lack of an adequate father figure, (*b*) dominant overprotection by the mother, (*c*) castration or castration fear, (*d*) frigidity, (*e*) rejection of pregnancy, and (*f*) the fact that the husband, an inadequate individual, leaves or threatens to leave the patient—play an important etiological part in the development of a phobic syndrome, and that such variables are not etiologically significant in the cases of anxiety neurosis and conversion hysteria which were used in ths study. One of the points, marriage to an older man, appears to be of no value whatsoever in differentiating among the groups.

The occurrence of these variables seems to be moderately systematized in the case of the phobia group, with lack of an adequate father figure occurring prior to dominant overprotection, which in turn precedes marriage to an older man. This is then followed by husband leaves, then by either frigidity or castration, and finally by rejection of pregnancy. In the case of the other two groups, no pattern of precedence of any type is shown.

Further, it is concluded that abstracts of case histories may be used instead of full case history data for making judgments of the type involved in the rating sheet designed for this study, thus effecting a major saving in time expended by the judge. The rating sheet appears to be reliable for use in further research when used in conjunction with the norms cited in the body of this study.

Treatment of the husband of the patient concurrently with the wife's own treatment seems to be the most efficacious manner of securing a remission of the phobic symptoms.

In addition to the points under consideration, it appears that phobic patients, in significantly greater proportion than patients in either of the other two groups, expose themselves sexually to the therapist. Over 90 per cent of both the husbands and fathers of the patients in the phobic group appear to have been psychopathic.

An implication drawn from this study is that time and effort may be saved by employing abstracts of case histories rather than the full history itself—that is, since the reliability between judgments is high, considering that full case histories were used in one instance and only abstracts in the second case, the abstracts are almost as useful in making judgments as the full material.

APPENDIX A

Illustrative Case Studies

Case 1

The patient is a white female, age 30. Her father, age 56, is a physician. Her mother is age 52. Her husband is age 32 and a minor executive of a railroad manufacturing concern. The patient and her husband have two children, both of whom are girls, ages 5½ years and 18 months. The patient was referred by her father after having visited a number of other psychiatrists. The presenting complaint was the inability of the patient to remain alone, both at home and primarily when away from home. Her greatest fear is being on the street alone. The patient is afraid of driving the car alone. She has no fear when her children, or her husband, or a relative is with her to drive the car or to go down town. Her present illness began during the first month of her first pregnancy. At this time she was very much afraid to be alone and, moreover, was "petrified" for fear of what would happen to her for having the baby. When the first baby was three months old, the patient's father and mother went on a vacation, and the patient felt that she was being "pushed out of the house."

It was during their absence that her symptoms became full-blown, and have remained for a little over six years now. The patient's mother states that her own life since marriage has not been a happy one because of her husband's drinking, gambling, and philandering. She stated that he was rarely home, and when home, he paid little or no attention to the child (the patient). On occasion he would bring her a magnificent present. The patient's father states that he was unable to remain in his home because of his wife's nagging. He stated that she spoiled and pampered the child, dominating her tremendously. The patient's husband states that the patient tries to boss him too much, nags him continually, and that their sex life is not satisfactory. He states that she "lies on the bed like a rock" when they have intercourse. He feels that her illness is either imaginary or a deliberate attempt to gain sympathy. Prior to the beginning of her illness, he threatened to divorce the patient unless she became a better wife. The patient's husband leaves the house about 7:00 o'clock in the morning and returns to the house about 6:30 in the evening. However, many evenings he is required to work until 10:00 or 11:00 o'clock. The "working late" began during the war and has continued until the present. The husband is quite resentful of the money which has been spent on the wife in seeing psychiatrists and will himself only agree to bring the wife to the office and to take her home. On no account will he enter into a therapeutic relationship himself. The husband admits that he has drunk quite a bit, has gambled to a small extent, and has not been faithful to his wife. He states that from the time he was about 13 until recently a maternal aunt has given him five dollars on frequent occasions to have intercourse with her. He was quite resentful of being interviewed in any case and seemed to have no regret for any of his past actions. Communication with the last psychiatrist whom the patient had seen revealed that he considered her to be "quite masculine in character and resentful of any attempt to deprive her of this trait." He stated that prior to her marriage the patient had indulged in perverse sexual activity, both homosexual and heterosexual in nature. Further, that she is quite resentful of her children and the restrictions which they place on her activities. He feels that her protest against her first pregnancy was one of the factors upon which the development of the illness is dependent. The physioneurological findings are negative except for a fistula between the vagina and the rectum. Consultations with surgeons reveal that although the fistula is not inoperable, it is unlikely that repair could be successful. The patient states that flatus escapes through the vaginal opening rather than through the rectum, causing her much discomfort. The patient has, on several occasions, exposed herself to the psychiatrist and the psychologist whom she is now seeing. However, she states that on several occasions when she has had an opportunity to go walking or riding with a man other than her husband or a member of her family, she has become most fearful. The patient states that she feels she is quite like her mother and hates herself for being this way. She is quite fond of her father. Several times there has been a remission in symptoms only to have them return in full strength. Whenever there does seem to be a remission in symptoms, the husband of the patient seems to "blow up" and tells the patient that she thinks only of herself. The patient states that she "gets scared" because he sounds like "mother." This is the first marriage for both the patient and her husband. They have been married eight years. Their religious faith is the same. The patient and her husband have sexual relations about once in two months. The socioeconomic status of the patient and her husband is above average. The formal diagnosis in this case is anxiety hysteria.

Case 2

The patient is a white female, age 33. Her husband is age 42. He is a clerk in a small combination beer garden and store. It is the patient's first marriage and the husband's third. The husband has no children from either of his previous marriages. The patient and her husband have no children. The patient had a bilateral salpingectomy at age 21. Her present illness began with nervousness six years ago and her presenting complaint was fear of being alone and particularly of going out on the street alone, which began two years ago. The intensity of this fear varies from time to time. She was able to work for about eight months after her "fear" symptoms began, and then had to quit her job because she was unable to go to work alone. When she attempts to go on the street alone, she complains of vertigo, syncope, palpitation, and perspiration. The patient is quite aggressive and states that she has fears of homosexuality. She feels these fears were partly related to homosexual experiences in which she participated during her adolescent years. She stated that in her early teens she hated men and decided never to get married, but that in her late teens she had several affairs with married men. They seemed more attractive to her because of their seriousness. She considered men of her own age as "silly and mushy." The patient was a telephone operator until she could no longer work because of her symptoms. The patient's father died when she was 17. The patient's mother has been living with them since their marriage. The patient states that her husband was quite intolerant of her illness and used to rant and rave because of it. She states that he is a drunkard and that she doesn't trust him. The patient's mother is very religious and quarrels with the patient and her husband because they go out and the mother will pout and isolate herself from the patient. The patient's mother states that she and the patient's father never got along well and that he was not faithful to her because he drank and because he gambled. The patient states that her husband also gambles. Her husband consented to bring her to a psychiatrist (having been referred by Johns Hopkins Clinic) only because he thought that after one visit she would be cured. They have many arguments about the need for a long-range program of treatment. Actually, she has spent over a thousand dollars on treatment and the husband is most resentful of this. He has threatened to leave on several occasions and storms out of the house and later comes back. The patient's husband says that she is too cool and too selfish. He calls her a Virginia Blueblood or Queen So-and-So. He is resentful of her spending money on clothes, particularly since she is tall and has to have most of her clothes made. He feels that he has to keep moving all the time. He states further that his wife is the "frigid beast." The patient states that the husband is a real "heller." The patient's husband states that he resents the mother-in-law living with them very much because he feels that she has more influence on the patient than he does. The patient feels that she must be two jumps ahead of everybody. She feels that people are not to be trusted. She feels very anxious. She does try to go out of the house and see the man on the street. On several occasions there has been a partial remission of symptoms, but as soon as the patient gets better, her husband becomes very aggressive toward her and she relapses. Both of them feel that they cannot discuss things without ending up in an argument. The patient began having homosexual and heterosexual relations at age 14. The patient states that her father was an alcoholic and was unfaithful to her mother. The patient's husband resents her knitting or reading or doing anything alone. The patient states, "I've always resented men deeply because of father's behavior." The patient becomes very nervous when she attempts to have intercourse with her husband and thinks that it is very unsatisfactory for her. She feels very resentful of her mother living with her and her husband. She feels that her mother has always tried to "boss" her too much and has pouted if she could not get her own way. Communication with the psychiatrist whom she saw at Johns Hopkins revealed that, "she seems to have confused her masculine and feminine identification and in many respects is somewhat masculine. She is resentful of male aggression." Her husband has no church affiliations and their socioeconomic status is below average, although the patient came from a wealthy family. The father of the patient was rarely at home before he died. The physioneurological examination is negative except for a low metabolic rate which is being corrected with thyroid extract. The formal diagnosis in this case is anxiety hysteria.

Case 3

The patient is a white female, age 30. The husband is age 33. It is the first marriage for both. They have one son, age 4. The patient's present illness began when her son was six months old. The boy became ill with a throat infection and the patient's husband was away at the time. She suddenly became nervous, began to tremble, and within a week her fear of being alone and going out on the street alone developed to an acute stage. The patient states that she had a great deal of anxiety when she was pregnant because she didn't know if she would be able to care for a baby. She had a very long labor and gained 15 pounds during the pregnancy. The patient's mother came to help her, but was highly nervous and more hindrance than help. The patient states that she never wanted a baby and did not want the responsibility of having one. The patient's mother stayed until the baby was two months old. The patient states that she has been completely frigid since her marriage, but was not so before her marriage. She states that she has no interest in sexual relations and that they rarely have them. When the patient was in her teens, she learned that her father had had a mistress since she was five years old. The patient's father told her that he wanted to have a good time and the patient's mother wanted a nice home and to raise her children; therefore, if he had a mistress, both he and his wife would be satisfied. The patient's father would usually come in late after the patient had gone to bed so that she rarely saw him. The patient says that she had more respect for her mother than for her father, although she was fond of her father and felt sorry for him because the patient's mother nagged him so much. She states that her mother nags her and acts very much like a martyr when the patient does not do what she wants. The patient's husband is the manager of a shoe store. He goes to New York on buying trips from three to four times a year. She has suspected him of being with another woman, but has never been able to prove it or to get him to admit it. As the patient got older, she realized that her parents didn't get along and thought that sex might have something to do with it, and was determined not to have the same trouble in her own marriage. Although she indulged in both homosexual and heterosexual play in her teens, she states that she was a virgin at the time of her marriage and that she knew more and had read more about sex than her husband. The patient's husband told her that she was abnormal because she had no orgasm. They began arguing shortly before the baby was born. The patient did not want to have a baby, nor had they planned to do so, although the husband wanted one. He states that perhaps it became too important to him. He has told her that he would like to go back into the army because he is unhappy at home. He says that while he was in the army, he lost his "niceness" for a while. He says that he did expect more of his wife after the baby was born, but got less. He says that she didn't seem glad to see him when he came home at night and that she spent a great deal of money. The patient states that the baby is too great a responsibility, that she never wanted the child and now she has to do all the baby's things by hand, get the groceries, and work around the house. The patient feels that their sexual relations can never work out and can never bring her happiness. She states that she is somewhat frigid now. She is fearful of being homosexual. She is quite fearful of her husband's being unfaithful. She wonders if her husband lies to her. Both the patient and her husband blame each other for their poor sexual relations. The patient's husband suffers from premature ejaculation. She feels that she has become almost totally frigid since the birth of the child. The patient's husband admits that he has not been an "angel" since the marriage. The patient's husband is quite resentful toward the patient's mother and does not like for her to visit them because he feels that the patient is both fearful of her and dominated by her. The patient admits that she is dominated by her mother to a great extent when she is around her. The patient says that she is somewhat fearful of her husband. She states that she feels aggressive toward both her husband and her mother and feels quite guilty and fearful because of it. She states that "if he (the husband) is unfaithful, it is because I have rejected him." She feels that his business trips have increased in the last few years and that he does this on purpose so that he can get away from her. The referring physician stated that he felt that the patient had confused her identification. She is particularly confused over her relationship with her father, particularly since the father was seldom home and was quite indulgent when he was home. The patient states now that she both hates and loves her father and that she wonders if it isn't the same way with her husband. Both the patient and her husband are perfectionistic and the patient has had some compulsion since she was in her early teens. For example, she cannot go to sleep at night unless her arms are underneath the covers and she awakens at night if she happens to toss and turn and one or both arms get out from under the covers. They have sexual relations infrequently at present. The physioneurological examination was negative. Originally present complaints of syncope, asthenia and anorexia, vertigo, palpitations, and cephalalgia continue, but to a somewhat less degree than originally. The socioeconomic status of the patient and her husband is about average.

APPENDIX B

Number the manifestations which are listed below in the order of chronological occurrence as you have noted it from your observation of the material provided to you. Place an (X) in the appropriate box indicating your judgment of the depth of the manifestation, and in that same row place the statement which leads you to the conclusion that such manifestation exists in noticeable degree. If there is more than one such statement which led you to your conclusions, enter all such portions of the material. In the last two rows, specify a personality type as psychopathic, neurotic (type), schizophrenic, etc., or "No Remark," using any personality type as being purely descriptive rather than feeling that you are committing yourself to a formal diagnosis, and adding any remarks you may wish concerning the personality of the husband and father.

Case No. ________

Number	Manifestation	Depth of Manifestation			
		Quite Marked	Moderate	Barely Noticeable	Not Present
	Lack of an adequate father figure				
	Dominant overprotection				
	Marriage to an older man				
	Castration or castration fear				
	Rejection of pregnancy				
	Frigidity				
	Husband leaves				
	Other				
Behavioral characteristics of husband		State:			
Behavioral characteristics of father		State:			

Vol. 67, No. 18 Whole No. 368, 1953

Psychological Monographs: General and Applied

Assessing Social-Emotional Climate in the Classroom by Withall's Technique[1]

Harold E. Mitzel and William Rabinowitz
Division of Teacher Education, Municipal Colleges of New York City

I. BACKGROUND OF THE STUDY

IN RECENT years educators have become increasingly concerned with psychological aspects of the classroom environment. It is now generally believed that the goals of education are not maximally achieved unless the social-emotional climate of the classroom is characterized by an atmosphere of warmth, mutual respect, and permissiveness. It is further assumed that the teacher is the most important individual in determining classroom climate, and that her verbal behavior is largely the medium for projecting her influence in the situation.[2]

Based upon these assumptions, Withall (14) has developed a technique for assessing social-emotional climate in the classroom through an analysis of the statements made by the teacher as she conducts classroom activities. It is possible, employing this technique, to place each statement made by a teacher into one of several categories.[3] Starting with an original list of twenty-five categories, Withall found it possible to reduce this number to seven which form a learner-centered to teacher-centered continuum. These seven categories cover all teacher statements in a parsimonious, yet psychologically meaningful fashion. The several categories, and brief examples of each, are as follows:

CATEGORY 1: *Learner-supportive statements* express agreement with ideas, actions, or opinions of the learner, or commend and reassure him. The dominant intent of these statements or questions is to praise, encourage, or bolster the learner. Statements of this kind frequently take the form of a single word response such as "Right," "Yes," or "Good."

CATEGORY 2: *Acceptant or clarifying state-*

---

[1] It is a pleasure for the authors to acknowledge the considerable assistance and support they have received while carrying out this study. Professor Withall, whose technique was employed, supplied us with sample transcripts and kindly encouraged our efforts. To Mrs. Pauline N. Cagen, Principal of Public School 127 in New York City, we are indebted for unreserved cooperation in providing the observational setting. Mrs. Dora J. Holt, our secretary, provided stenographic assistance in the classrooms and later persevered, with undisturbed good humor, through successive revisions of the manuscript. To the four teachers who were observed in this study we owe a special debt of gratitude. Being observed necessarily entails certain irritations and frustrations. These they suffered without complaint. The warmth of their reception did much to make this study an enriching personal experience for the authors.

[2] Some recent practice (8, 16) favors the use of the feminine pronoun in referring to teachers, and since only women teachers participated in this study, the feminine gender seems quite appropriate.

[3] A statement is defined in this report as any type of verbalization made by a teacher. Statements require no particular grammatical construction and may consist of questions, declarative sentences, exclamations, and imperatives. In a personal communication from John Withall, dated December 22, 1952, statements are further defined: "The criterion of 'a statement' is the expression of a complete idea or concept . . . any given remark of a teacher may be broken up into two or more statements depending on how many discrete ideas are expressed in said remark."

ments show considerable understanding by the teacher or restate clearly and succinctly in the teacher's words the ideational or emotional content of the learner's statements. The dominant intent of these statements is to help the learner gain insight. Statements of this kind are rather infrequent, but the following is an example: "In other words you think that if we decide to go back again we should plan to spend the entire day."

CATEGORY 3: *Problem-structuring statements* offer facts, ideas, or opinions to the learner about phenomena or procedures in a nonthreatening objective manner. The dominant intent of these statements is to sustain the learner by facilitating his problem-solving activities. A good example of this kind of statement is the following: "What is special about twins? Two brothers are born on the same day one year apart; are they twins?"

CATEGORY 4: *Neutral statements* evidence no supportive intent, being neither teacher-centered, learner-centered, nor problem-centered. The dominant intent of these statements is to facilitate the administrative and procedural detail of class management. Examples of neutral statements are the following: "Children, there will be no movie today since our school projector is broken." "Who is going to march in the St. Patrick's Day parade tomorrow?"

CATEGORY 5: *Directive statements* advise the learner regarding a course of action and offer him no choice or severely limit his choices. The dominant intent of these responses is to have the learner take up the teacher's viewpoint and pursue a particular course of action. An example of a directive statement is the following: "If there are any words which you do not know, put them down in your notebooks and we will go over them tomorrow."

CATEGORY 6: *Reproving, disapproving, or disparaging statements* express complete or partial disapproval of the ideas, behavior, or personal characteristics of the learner. The dominant intent of these statements is to admonish the learner for unacceptable behavior; to instruct the learner in societal values as interpreted by the teacher; or to impress upon the learner that he has failed to reach certain standards of excellence. A good example of mild reproof is the following: "Children, you think you are whispering, but you are not." Stronger reproof is found in this statement: "You knew that yesterday, but now you've forgotten it completely."

CATEGORY 7: *Teacher-supportive statements* are those in which the teacher refers to herself, expressing a defensive attitude, or refers to her present or past interests, activities, or possessions in order to reassure herself and confirm her position. The dominant intent of these teacher responses is to assert, defend, or justify the teacher. An example of this type of statement is the following: "You must say 'excuse me' when you interrupt me; I say 'excuse me' when I interrupt you. Now we will go on without any interruptions."

It is important to note that in using the Withall procedure one categorizes statements not in terms of their *content* but according to the *intent* of the teacher who makes them. Theoretically, the verbal behavior of the teacher, as assessed by the Withall procedure, is uninfluenced by the subject matter taught. In practice, however, it seems likely that such content factors as subject matter do exercise some influence on the frequency of verbal statements in certain categories. The extent of this influence, if any, is not known.

The intensity of a statement is not a variable in the Withall procedure. The technique provides a numerical measure of statements in each category, but the manner in which these statements are made is not assessed. Statements made with great emotion are weighted equally with those made with serene calm or disinterest. This characteristic does not appear to represent a profound problem or limitation, since it is reasonable to expect that intensity of statements and frequency of statements are closely related.

A procedure for categorizing statements necessarily ignores several other facets of communication among human beings. A teacher's posture, gestures, and facial expressions undoubtedly convey meaning to children, and therefore influence the social-emotional climate of the classroom. Even silence may be put to effective use by teachers in the control of classroom situations. Other characteristics of teachers' verbal behavior which conceivably influence classroom climate are the rate and clarity of expression. An essentially dominative social-emotional climate may be maintained by a teacher who speaks so rapidly and indistinctly that the pupils are insecure in her presence. Though these variables are important factors in the communication between teacher and pupil, they are extremely difficult to assess objectively. Moreover, on logical grounds it would seem that measures of social-emotional climate based upon the categorization of the statements of teachers should be highly correlated with other measures emphasizing different aspects of the verbal pattern.

In his original development of the Climate Index, Withall (14) worked with typescripts made from sound recordings of regular sessions of several high school classes in the University of Chicago Laboratory School. The Index was defined as the ratio of the number of learner-centered statements (sum of Categories 1, 2, and 3) to the number of teacher-centered statements (sum of Categories 5, 6, and 7). Objectivity of the Index was determined upon the basis of agreement among several judges classifying statements in the typescripts. An estimate of validity was obtained by comparing the rank order of three typescripts on the Climate Index with the rank order of Anderson's Integrative-Dominative Ratio (1) using the same three typescripts. Other evidences of validity include a questionnaire evaluation by a class of 25 pupils and the use of several class sessions in which the climate was experimentally varied by the teacher. Reliability was studied by comparing successive blocks of 50 statements from the typescripts of one class session. The extent to which changes occurred in the proportions of statements in each category was noted. Proportions based on the first 200 statements did not significantly change as successive blocks of statements were added. It was therefore concluded that 200 statements offered a reliable sample of a teacher's classroom verbal behavior.

Flanders (5) showed that the Climate Index, when used in an experimental situation with one teacher and one pupil, was related to such physical manifestations of anxiety as galvanic skin response and pulse rate. Using the Index as a criterion of group climate, Perkins (10) categorized the statements of six leaders of in-service teacher training groups. He found that changes in group climate were paralleled, to a marked degree, by changes in group learning.

A further study by Withall (16) had as its purpose the assessment of the classroom climates experienced by a seventh-year class in the Laboratory School as it changed from teacher to teacher. This study emphasized the fact that different teachers create quite different psychological atmospheres for the same group of youngsters. An important change in the procedure for categorizing statements was adopted in this study. Instead of recording class sessions and making categorizations from typescripts, the observer classified teachers' statements in the "live" situation, i.e., as they were made in the classroom.

Thus it can be seen that since its original development, additional research experience has been gained with the Withall technique. Applications have ranged from junior high school classrooms to adult groups enrolled for in-service teacher training, from a "class" involving a single pupil to classes composed of large numbers of pupils. Although used exclusively in educational settings thus far, the method undoubtedly is applicable in the study of psychotherapy, counseling, group discussion, industrial interviewing, and other human relationships in which interaction takes place mainly on a verbal level.

II. PROBLEM

The review of the literature cited above indicated the desirability of answering certain basic questions concerning the usefulness of a verbal-behavior method in studies of teacher effectiveness. Withall's technique seemed suitable as a predictor or criterion of teacher effectiveness (9, 11), but its immediate use in this capacity appeared to depend upon its reliability. Accordingly, a study was planned which might answer three fundamental questions, usually subsumed under the general heading of reliability. These questions are as follows:

a. Can self-trained observers, working independently and observing the same behavior, agree in their classification of teachers' statements?

b. Is the verbal behavior of teachers stable from one classroom visit to another?

c. In a single school, is the categorization of verbal behavior sufficiently sensitive to reveal differences in the classroom climates provided by different teachers?

A satisfactory answer to each of these questions seems preliminary to the design of studies of teacher effectiveness employing the technique.

III. PROCEDURE

Exploratory Efforts

The initial efforts of the authors were directed toward a clarification of the meaning of the various categories. It was essential that the authors have a common understanding concerning the types of teacher statements which properly fit into each category. Professor Withall, in a private communication, encouraged the authors to employ the method and provided them with several sample typescripts as practice material. The principal of a small Manhattan elementary school, with the consent and approval of her 13 faculty members, agreed to furnish the observational situation and to make minor adjustments in the school schedule for the convenience of the observers. At aperiodic sessions during a six-week interval, classroom visitation and practice with the classification scheme were undertaken. It soon became apparent that sound reproductions of class sessions using commercial tape recorders were not feasible because of the high noise concentration in the classrooms. Stenographic notes of teachers' verbalizations were found to be more satisfactory than tape recordings, but even with this method the authors estimated that no more than 75 per cent of a teacher's statements were obtained during any given observation period. After about 20 clock hours of experience in classifying classroom verbal behavior of teachers from typescripts (see Appendix), the writers attempted to dispense with reproduction procedures and to categorize statements in the "live" situation. As a teacher conducted a class, the observers (the writers) independently categorized her statements. Practice for a few sessions with this method seemed to yield close agreement between the observers as to the total number of statements made by the teacher and the proportionate numbers of statements in the several categories.

Observational Setting

The school in which the observations were made is somewhat atypical for the country as a whole. It is a small, six-year elementary school, housed in a 100-year-old building, and located in an economically underprivileged area of New York City. Approximately 30 children are enrolled in each class. Many of the pupils are foreign born; about one-half are natives of Puerto Rico. Most of the children are bilingual. Measured IQ's are almost all below 100, but this is certainly attributable, in part, to difficulties in the use of the English language.

From an impressionistic point of view the atmosphere of most classrooms in this school was free and informal. The authority of the teacher, though expressed in subtle rather than obvious ways, was seldom challenged. Disciplinary problems were infrequent but not unknown. The teachers, if one may characterize a group of teachers in terms such as these, are neither "traditional" nor "modern" but at some indeterminate point between these two extremes. Such practices as severe regimentation and repetitive drill were never seen; neither was there evidence of the studied encouragement of pupil self-direction.

It is impossible to estimate the extent to which these teachers are representative of teachers in general. Like most New York City teachers, they are well educated, with a strong, professional interest in their work. Their original attitude toward the study could best be characterized as "cautious toleration." During the course of the observations, however, this attitude underwent a marked change. Rapport was established as the observers became more familiar and less threatening figures. The teachers gave definite evidence of enthusiastic interest and unreserved cooperation as the study progressed.

Design

Three important aspects of the categorization of teachers' verbal behavior were the principal objects of study: (*a*) differences between observers, (*b*) differences among observation periods or visits, and (*c*) differences among teachers. A 2 × 4 × 4 multiple-classification design involving two observers, four visits, and four teachers was employed. This was considered a first series of observations. It was followed by a second series of observations, of identical design, in which the same observers visited the same teachers on four more occasions. Dividing the eight observation days into two separately analyzed series of four visits each enabled the observers, after the first series, to re-evaluate their criteria for classifying statements into the several categories. It was thought that with this method the categorizations by observers could perhaps be brought into closer alignment for the second series of four visits.

The data were to be treated by the analysis of variance method, and each category of statements was to be analyzed separately. During the exploratory work, however, it became apparent that statements in Category 2 (acceptant or clarifying) and Category 7 (teacher-supportive) were very rare. It was therefore decided that analysis should be restricted to each of the five "most tallied" categories and a summation category (Climate Index).

Two fourth-year and two fifth-year teachers in the cooperating school were chosen for study. None of the classrooms of these teachers had been visited during the practice and exploratory sessions, and the observers had no knowledge of the methods and techniques employed by these teachers prior to the first observation. Observations were held on Monday mornings for eight consecutive weeks, exclusive of school holidays. Each teacher was visited some time during the morning. The order of visitation of classrooms can best be described as haphazard. Since time of day may conceivably be related to classroom climate, it was desirable to avoid a systematic order. At the same time it was impossible, because of scheduling difficulties, to utilize a randomizing technique. On two occasions, as a result of absence, it was not possible to visit a particular teacher on Monday morning, and a later day in the same week was chosen as a substitute.

Previous experience with Withall's procedure suggested that approximately 100 statements would provide a usable sample of a teacher's verbal behavior for a single observation period. Withall (15) concluded that a sample of 50 statements was sufficient for bringing out gross differences in classroom climate and that 200 statements were sufficient for estimating a teacher's position along the climate continuum. However, since it seemed desirable for the two observers to impose upon a given teacher no longer than 30 minutes per day, a goal of approximately 100 teacher statements per visit was selected. In practice the number of statements tallied by the two observers on the 32 visits of both series varied from 74 to 133 per visit, with a median of 100 statements. Data relative to this point are presented in Tables 1 and 2.

Factors not studied. Ultimately, extensive evidence on the validity of Withall's procedure for categorizing teacher statements will have to be obtained. The pupil-growth correlates of a teacher's verbal behavior must be sought. The discovery of such correlates, however, is not within the scope of this study, nor does this study deal with an assessment, inference, or assumption concerning the effectiveness of the teachers who served as subjects.

It should be noted that several factors which may ultimately prove to influence observed classroom climate have been uncontrolled in this study. Factors of this type would include the day of the week, the time of day, and the background and experiences of the pupils. It was the feeling of the writers that, though these factors are worth studying, they contribute a small proportion to the variation in observed classroom climate.

Assumptions. Several assumptions are required for the multiple-classification design and analysis of variance technique to be strictly applicable. Cochran (3) has set forth these assumptions as independence, homogeneity of variance, and normality of distribution of the experimental errors, as well as additivity of treatment effects. Concerning this study, which was to involve the repeated observation of the same teachers by the same observers, it seemed that the independence of errors was most in jeopardy. To minimize this difficulty, special care was taken by the observers during each series of visits to avoid any exchange of impressions about the teachers. Also, by placing observation records in a file and by carefully refraining from totaling the tally marks in each category following an observation period, each observer was able in a measure to prevent contamination of his subsequent categorizations of statements. Since inspection of the data provided no evidence for heterogeneity, the assumption of homogeneous error variance was considered satisfied without rigorous test.

Nature of the variate. Following the precedent set by Withall (15) in his original development of the Social-Emotional Climate Index, it was decided to make the entry for each cell of the design the percentage of the total number of a teacher's statements falling in a given category. This procedure has the effect of equating all observation periods to a base of 100 teacher statements, making it possible to compare varying numbers of statements recorded by different observers during different visits. Distributions of percentages, however, are characteristically rectangular in form and transformation is required to normalize these values. Bartlett (2) points out the usefulness of the angular transformation for percentages and suggests the employment of Table XII in Fisher and Yates's *Statistical Tables* (4). This method was followed for these data. For example, if Teacher A has a total of 99 statements attributed to her by Observer 1 during a given visit, and 25 of these statements have been classified by him as "learner-supportive" (Withall's Category 1), then 25.3 per cent of her verbal behavior is placed in the first category. Using the angular transformation table, 25.3 per cent is transformed into an angle of 30 degrees, rounded to the nearest whole degree. In this way seven transformed values, one for each statement category, were obtained for each of four teachers, by each of two observers, on each of four visits, in two separate series of observations.

As anticipated, examination of the tables of transformed values indicated that it would be worth while to analyze only five of the seven statement categories and a Climate Index based on the sum of the first three categories. So few teacher statements were classified by the observers as belonging in Category 2 and Category 7 that an analysis of these was not undertaken.

The Index used in this study was the sum of statements in Categories 1, 2, and 3 (learner-centered statements) divided by the total number of statements. This Index was computed separately for each visit to each teacher by each observer. There were therefore 32 indices computed for each series of observations. It should be noted that the Index used in this study, since it is based upon proportions of the total number of statements, will vary between zero and unity (before transformation). Withall's ratio of learner-centered statements to teacher-centered statements, ignoring neutral statements (Category 4), will vary between zero and some indeterminate value, depending upon the number of statements categorized by the observer.

IV. ANALYSIS OF RESULTS

Comparisons of Numbers of Statements Recorded by Observers

One of the preliminary questions for study in this project was, "Can two persons, observing simultaneously and independently in the classroom, agree upon the total number of statements made by a teacher?" Table 1 provides information relevant to this question. During the first series of observations there was some tendency for Observer 2 to record more statements than Observer 1–1,741 compared with 1,693. Differences between observers on number of tallied statements for the 16 pairs of observations (four visits to four teachers) resulted in a t value of 2.216, which is significant at the 5 per cent level ($t_{.05}=2.131$; 15 df). Thus, during the first series of observations one observer tended to record a significantly larger number of statements than the other. This difference, although undesirable, does not necessarily lead to disagreement between observers when statements in individual categories are analyzed. Statements were analyzed for each observer in the separate categories by calculating a *ratio* based upon the total number of statements tallied during the observation period. Two ratios may be identical though the total number of statements on which they are based differ markedly.

Between the first and second series of observations the observers redefined their criteria of a statement in order to reach greater agreement. The success of their effort is reflected in Table 2. The mean difference for the 16 pairs of observations of this second series resulted in a t value of 0.260, which is nonsignificant. Thus, during the second series of observations, after the observers re-evaluated their criteria, the systematic significant difference in the number of recorded statements was eliminated.

Comparison of Range and Median Climate Indices

Table 3 presents the Social-Emotional Climate Indices of the four teachers who

TABLE 1

TOTAL NUMBER OF STATEMENTS AND LENGTH OF RECORDED OBSERVATION PERIODS

First Series

		Teacher									
		A		B		C		D		Totals	
Visit	Obs.	*N*	Time	*N*	Time	*N*	Time	*N*	Time	*N*	Time
I	1	105	20′	117	16′	105	20′	137	20′	464	76′
	2	109		126		116		138		489	
II	1	95	24′	124	13′	109	23′	89	12′	417	72′
	2	93		123		111		104		431	
III	1	78	21′	106	24′	110	30′	102	17′	396	92′
	2	76		101		111		107		395	
IV	1	93	18′	123	22′	84	27′	116	22′	416	89′
	2	96		120		87		123		426	
Totals	1	371	83′	470	75′	408	100′	444	71′	1693	329′
	2	374		470		425		472		1741	
Medians	1	94	20.5′	120	19.0′	107	25.0′	109	18.5′		
	2	95		122		111		115			

TABLE 2

TOTAL NUMBER OF STATEMENTS AND LENGTH OF RECORDED OBSERVATION PERIODS

Second Series

		Teacher									
		A		B		C		D		Totals	
Visit	Obs.	*N*	Time	*N*	Time	*N*	Time	*N*	Time	*N*	Time
I	1	99	16′	95	11′	74	25′	88	15′	356	67′
	2	100		92		79		93		364	
II	1	93	25′	87	19′	92	30′	87	28′	359	102′
	2	85		85		95		85		350	
III	1	120	23′	133	23′	91	35′	100	18′	444	99′
	2	108		132		89		106		435	
IV	1	101	20′	78	20′	94	32′	92	22′	365	94′
	2	102		86		104		89		381	
Totals	1	413	84′	393	73′	351	122′	367	83′	1524	362′
	2	395		395		367		373		1530	
Medians	1	100	21.5′	91	19.5′	91	31.0′	90	20.0′		
	2	101		89		92		91			

participated in the study for all observation sessions. A separate analysis of the observations made in the two series is not provided at this point, since that is considered in other sections of this report. The tabular values are based upon the following formula:

$$\frac{\text{Sum of Statements in Categories 1, 2, and 3}}{\text{All Recorded Statements}} \times 100$$

The three categories of the numerator (learner-supportive, acceptant or clarifying, and problem-structuring) may be considered learner-centered statements. Their ratio to the total number of statements made by the teacher is the classroom Climate Index employed in this study. Indices for the two observers were combined and averaged, since differences between the observers were generally not significant, as subsequent analysis will demonstrate.

Examination of the median values near the bottom of Table 3 shows that Teacher A, with almost 60 per cent of her statements in the learner-centered categories, has the highest Climate Index. Teacher C, on the other hand, has the lowest Index, 33.9 per cent. Teachers B and D have nearly equal Climate Indices, based on slightly less than one-half of their statements in the learner-centered categories. It was the subjective impression of the observers, prior to the analysis of the data, that the four teachers tended to provide similar classroom climates. A larger sampling of the teaching profession would undoubtedly yield a wider range of individual differences in average Climate Indices.

TABLE 3

SOCIAL-EMOTIONAL CLIMATE INDICES (PERCENTAGES) FOR FOUR TEACHERS ON EIGHT VISITS

Visit	Teacher			
	A	B	C	D
1	58.4	26.5	44.5	47.6
2	61.1	53.9	35.0	33.9
3	42.9	35.3	37.1	56.5
4	82.0	47.0	25.8	40.6
5	54.8	65.3	43.1	55.9
6	74.6	41.3	32.7	50.0
7	48.2	40.4	24.5	27.7
8	78.3	66.5	31.3	41.4
Median	59.8	44.2	33.9	44.5
Range	39.1	40.0	20.0	28.2

Table 3 not only shows differences among teachers in typical classroom climate but differences in the climate provided by the same teachers on different occasions. A picture of the variability in the climates characteristic of the four teachers is obtained by considering the range of percentage points covered by the Indices of each teacher for the eight observations. Teachers A and B are of approximately equal variability and are relatively more variable than Teachers C and D. These data suggest that not only do teachers differ among themselves on their Climate Indices, but they also differ in the consistency of classroom climate which they provide from occasion to occasion. These impressions are gained from an inspection of Table 3 and will be analyzed more thoroughly in the material to follow.

Results and Interpretation of Analyses of Variance

Table 4 shows the structure of the study for the first series of four observations, and Table 5 is a similar presentation for the second series of four visits. In all cases the second-order interaction ($T \times O \times V$) was used as the error term for testing lower-order effects. Results are shown separately for five categories of teacher statements and the Climate Index based on the sum of the statements in the first three categories.

The interpretations suggested in this

TABLE 4

SUMMARY OF ANALYSES OF VARIANCE FOR FIVE CATEGORIES OF STATEMENTS AND AN INDEX OF TEACHERS' VERBAL BEHAVIOR

First Series

	Source of Variation							
	Main Effects			Interactions				Total Mean Square
				1st Order			2nd Order	
	Among Teachers	Among Visits	Between Observers	T×O	T×V	O×V	T×O×V (Error)	
Degrees of freedom	3	3	1	3	9	3	9	31
Category 1 Learner-supportive								
Mean Square..	27.542	29.708	0.125	2.875	33.792	7.708	1.236	16.738
F............	22.28**	24.04**	<1.00	2.33	27.34**	6.24*	—	—
Category 3 Problem-structuring								
Mean Square..	301.135	3.125	6.125	2.458	91.347	0.458	4.458	57.738
F............	67.55**	<1.00	1.37	<1.00	20.49**	<1.00	—	—
Category 4 Neutral								
Mean Square..	7.865	82.865	22.781	17.531	29.670	16.365	3.448	22.410
F............	2.28	24.03**	6.61*	5.08*	8.60**	4.75*	—	—
Category 5 Directive								
Mean Square..	339.583	44.250	4.500	1.917	99.000	6.417	2.833	67.661
F............	119.87**	15.62**	1.59	<1.00	34.95**	2.27	—	—
Category 6 Reproving								
Mean Square..	38.458	48.875	6.125	1.125	29.431	0.708	1.819	17.899
F............	21.14**	26.87**	3.37	<1.00	16.18**	<1.00	—	—
Climate Index								
Mean Square..	348.792	22.042	3.125	4.792	127.819	5.375	4.597	75.415
F............	75.87**	4.79*	<1.00	1.04	27.80**	1.17	—	—
Tabled *F* Values								
$F_{.01}$..........	6.99	6.99	10.56	6.99	5.35	6.99		
$F_{.05}$..........	3.86	3.86	5.12	3.86	3.18	3.86		

* Exceeds the 5 per cent level of significance.
** Exceeds the 1 per cent level of significance.

presentation and discussion of results are characteristic of the four teachers studied and can be generalized only to whatever hypothetical population they represent. As in much educational research, it was not possible to select the particular elementary school at random from a population of New York City schools. Nor was it possible, for administrative reasons, to select randomly four teachers from among the particular school's 13 faculty members. There is, therefore, no reason to believe that the four teachers in this investigation are representative of teachers generally. In the future, if it is desirable to estimate parameters for the Climate Index or to construct norms, considerable attention will have to be given to problems of sampling.

As in the case of the teachers, the observers also constitute a sample from a hypothetical population of observers. Although the categories are objectively defined, it is possible that observers with radically different orientations might categorize the same teachers' statements in somewhat different ways.

Main effects. Of the three main effects which are of concern in this study, observers, visits, and teachers, the source of variation due to the observers is of initial importance. The third column in Tables 4 and 5 presents the between-observers mean square and appropriate *F* test for

TABLE 5

SUMMARY OF ANALYSES OF VARIANCE FOR FIVE CATEGORIES OF STATEMENTS AND AN INDEX OF TEACHERS' VERBAL BEHAVIOR

Second Series

	Source of Variation							
	Main Effects			Interactions				Total Mean Square
				1st Order			2nd Order	
	Among Teachers	Among Visits	Between Observers	T×O	T×V	O×V	T×O×V (error)	
Degrees of freedom	3	3	1	3	9	3	9	31
Category 1 Learner-supportive								
Mean Square..	69.615	181.031	0.281	0.281	17.920	1.365	0.698	29.830
F............	99.73**	259.36**	<1.00	<1.00	25.67**	1.96	—	—
Category 3 Problem-structuring								
Mean Square	679.781	46.615	2.531	2.281	116.976	2.615	2.698	105.596
F............	251.96**	17.28**	<1.00	<1.00	43.36**	<1.00	—	—
Category 4 Neutral								
Mean Square..	52.375	28.208	24.500	8.083	36.125	12.750	7.000	23.125
F............	7.48**	4.03*	3.50	1.15	5.16*	1.82	—	—
Category 5 Directive								
Mean Square..	664.365	186.781	1.531	1.031	115.781	3.615	2.226	117.128
F............	298.46**	83.91**	<1.00	<1.00	52.01**	1.62	—	—
Category 6 Reproving								
Mean Square..	76.115	28.698	13.781	0.365	44.003	1.115	3.809	24.612
F............	19.98**	7.53**	3.62	<1.00	11.55**	<1.00	—	—
Climate Index								
Mean Square..	498.531	236.948	3.781	1.865	70.920	1.281	2.920	93.039
F............	170.73**	81.15**	1.29	<1.00	24.29**	<1.00	—	—
Tabled *F* Values								
$F_{.01}$..........	6.99	6.99	10.56	6.99	5.35	6.99		
$F_{.05}$..........	3.86	3.86	5.12	3.86	3.18	3.86		

* Exceeds the 5 per cent level of significance.
** Exceeds the 1 per cent level of significance.

each category of teachers' statements in both series of observations. In the first series of visits differences between the two observers were attributable to chance variation in all categories except neutral statements, Category 4. Here the issue remained in doubt since the *F* test yielded significance between the 5 and 1 per cent levels. An examination of the means showed that Observer 1, on the average, placed 15 per cent of all teachers' statements in this category, while Observer 2 similarly categorized 13 per cent from the same set of teachers and visits. In the analysis of the second series of visits, shown in Table 5, the main effect due to observers is nonsignificant in all categories, including the category for neutral statements.

The remaining main effects, among-visits and among-teachers, are interesting, but were not the chief concern of the investigation. Most of the analyses reported in Tables 4 and 5 show highly significant results for the among-teachers and the among-visits sources of variation. These results indicate that random sampling variation fails to account for the observed variability in these sources. It should be emphasized, however, that we cannot say anything further regarding a population of visits or a population of teachers. The existence of a significant interaction between teachers and visits

(T × V), interpreted as being characteristic of the population, prevents us from making inferences regarding the "real" effects of either teachers or visits considered separately. At the same time we are prevented from using the teacher-by-visit interaction as an error term. Since it is significant, it does not provide an appropriate estimate of error; consequently, no meaningful null hypothesis is available to test. In nonstatistical language, the variation among teachers on different visits is of such magnitude that the variability among either teachers or visits is not made apparent for the majority of statement categories. This point will be explored further in the discussion of the interactions of the analysis of variance tables.

Interactions. Considering the first-order interactions listed in Tables 4 and 5, the general picture is one of nonsignificant effects for those interactions involving observers. The teacher-by-visit interaction with its nine degrees of freedom is in a sense the primary focus of the study of interactions. As indicated in the discussion of the main effects, this term (column 5 of Tables 4 and 5) was found to be significant for all of the categories on which the analysis was made. It shows that teachers are quite variable from time to time with respect to the proportions of their verbal behavior occurring in the categories of Withall's technique.

The extent to which differences among teachers, in general the largest single source of variation in the data, are masked by the variation due to the interaction of teacher-by-visit is an important problem. A rule-of-thumb procedure for examining this point consists of the formation of a ratio, the among-teachers mean square over the teacher-by-visit interaction mean square. Lindquist (7, p. 214) suggests that this ratio of variances is seldom amenable to the F test for significance because of the inability to postulate a meaningful hypothesis. The ratio does, however, when it is several times unity, provide an accurate notion of the stability of the respective rank orders of the elements being studied, in this case teachers. An example from Table 4 will make this point clearer. The ratio of the among-teachers mean square and the teacher-by-visit mean square for the Climate Index in the first series of observations is:

$$348.792/127.819 = 2.73$$

This ratio is not very large, and we would therefore not expect great stability in the rank orders of the four teachers over the four visits. If the data for the two observers are combined, the set of rank orders is given in Table 6. Using a nonparametric method recommended by Kendall (6), a coefficient of concordance computed from the ranked data in Table 6 was found to be .475. This value is nonsignificant, and it would therefore appear that the distribution of ranks in the Climate Index for the first series of observations might easily have arisen by chance.

TABLE 6

TEACHERS' RANK ORDERS ON CLIMATE INDEX FOR FOUR CLASS VISITS

First Series

Teacher	Visit			
	I	II	III	IV
A	1	1	2	1
B	4	2	4	2
C	3	3	3	4
D	2	4	1	3

An identical analysis for the second series of observations yielded somewhat different results. The ratio of the among-teachers mean square and the teacher-by-visit mean square for the Climate Index for this second series of observations is:

$$498.531/70.920 = 7.03$$

This ratio is much larger than unity, and, in addition, much larger than the comparable ratio for the first series of observations. If the data are combined for the two observers on the second series of visits, the set of rank orders of the four teachers on each of four visits is given in Table 7. It can be seen that the agreement among this set of ranks is quite close and that the average order for the teachers would be A, B, D, and C. The coefficient of concordance was found to be .700. The likelihood that the arrangement of ranks presented in Table 7 could occur by chance if there were no fundamental agreement among them is less than two in one hundred. Thus it appears possible, in spite of fluctuations from occasion to occasion, to order the four teachers participating in this study with respect to their standing on the Climate Index. The stability of ranks necessary for ordering the teachers did not become apparent, however, until the second series of observations. These findings, if they can be duplicated in other school situations among other teachers, suggest that the Climate Index represents a stable variable in spite of fluctuations from time to time for the same teacher. As in this study, stability of ranks may not become apparent until a large number of observations have been made. In school situations where teachers differ markedly, fewer observations will be required in order to demonstrate stability among ranks.

TABLE 7

TEACHERS' RANK ORDERS ON CLIMATE INDEX FOR FOUR CLASS VISITS

Second Series

Teacher	Visit			
	I	II	III	IV
A	3	1	1	1
B	1	3	2	2
C	4	4	4	4
D	2	2	3	3

Comparison of the First and Second Series of Observations

In general, although the two series of observations were comparable in virtually every respect, certain unanticipated differences between them appeared. The results reported in Table 5 for the second series of visits tended to confirm the outcomes of the first series reported in Table 4, i.e., significance tended to be characteristic of the same main effects and interactions. Examination of the two tables does, however, make it possible to suggest additional outcomes. There was a general tendency for the total variances of the six categories (last column) to be larger for the second series of observations. Along with this general increase in variability there was a tendency for the error terms to be smaller in the second series than in the first, with the exception of Categories 4 and 6. These effects corroborated two subjective notions entertained by the observers:

a. Practice effects for making categorizations of teachers' statements were believed to have occurred. The observers felt that, with continued practice and the discussions between the first and second series of observations, they became more precise in their ability to recognize and categorize statements. Increased skill in the recognition and categorization of teachers' statements seems to offer a reasonable explanation for the slight decreases in most error terms from the first to the second series.

b. Teachers' verbal behavior was markedly influenced by the presence of the observers in the classroom. Initially, the teachers seemed to show an undue awareness of the observers. The observers got the impression, during the first few visits, that they were witnessing "tailor-made" lessons. Both the content and the manner of presentation seemed to have been especially prepared for the observers' benefit. As repeated visits were made to the same classrooms, it was felt that this rather obvious planning became much less important. During the later visits the teachers seemed to provide a more typical, less guarded classroom performance. On logical grounds, it appears likely that the net effect of "preparation for being observed" is to reduce the differences among teachers in all statement categories and to have a general depressing effect upon the total variance for each category. Though it cannot be proved conclusively, this is a plausible explanation for the increase in the among-teachers mean square and in the total variances from the first series of visits to the second.

V. SUMMARY AND CONCLUSIONS

This study was designed to investigate three aspects of the reliability of teachers' verbal behavior. The specific questions on which the project was focused are as follows:

a. Can self-trained observers, working independently and observing the same behavior, agree in their classification of teachers' statements?

b. Is the verbal behavior of teachers stable from one classroom visit to another?

c. In a single school, is the categorization of verbal behavior sufficiently sensitive to reveal differences in the classroom climates provided by different teachers?

The method developed by Withall (13, 14, 15) for assessing classroom climate by classifying verbal behavior along a learner-centered to teacher-centered continuum was employed. Using Withall's seven categories for tallying teachers' statements, a multiple-classification design was developed in which two observers (the writers) visited the classrooms of four teachers in an elementary school on four different occasions approximately one week apart. This design constituted the first series of observations and was followed by a second series of identical design involving the same teachers and observers.

In analyzing the data the first problem considered was the extent to which the two observers agreed on the total number of statements uttered by the teachers during each observation. In the first series of 16 observations a statistically significant difference (5 per cent level) between the observers was found. One observer tended to record a slightly larger number of statements than the other. Between the first and second series of observations the two observers attempted to reach greater agreement on the defining characteristics of a statement. This effort was apparently successful, since no significant difference in the number of recorded statements was found on the second series of observations.

A Climate Index based upon the proportion of learner-centered statements to all recorded statements was calculated for the four teachers on the eight observation sessions. A descriptive presentation of these indices indicated that teachers vary in the typical classroom climate they provide. In addition, teachers appeared to differ in the consistency of classroom climate which they provide from one occasion to another.

The data were further treated by

means of the analysis of variance, five statement categories and the Climate Index being analyzed separately. The most important outcomes of this analysis were:

a. In the first series of visits differences between the two observers were attributable to chance variations in all categories except neutral statements, Category 4. In the second series of visits, however, the main effect due to observers was nonsignificant in all categories, including the category for neutral statements.

b. The factors accounted for by differences among the observed teachers and among visits were attributable to nonrandom sources. Since the teacher-by-visit interaction term was found to be significant, no further interpretation regarding these main effects could be made.

c. The significant effect for teacher-by-visit is an answer to the questions posed in this study concerning differences among teachers and stability of teachers' verbal behavior from time to time. It was concluded that teachers showed marked fluctuations in their verbal behavior; "real" differences among teachers were apparent in spite of these fluctuations. It was not until the second series of observations, however, that the teacher differences were clearly manifest.

In the comparison of the results of the first and second series of observations, two impressions stood out. The observers were able to increase their skill in categorizing teachers' statements during the course of the experiment, and teachers' verbal behavior was markedly influenced by the presence of the observers in the classroom, particularly during the early visits.

Concerning the generalizability of these findings it should be noted that the observed teachers are not representative of teachers in general. They were elementary teachers in a particular school situation in the largest city in the nation. Had it been the writers' purpose to estimate the typical classroom climates provided by teachers in the country at large, then the data would be seriously inadequate. Since this was not the goal, the fact of unrepresentativeness is less damaging. A second qualification involves the fact that samples of approximately 100 teacher statements were tallied by the observers during any given visit. There is no way of inferring from the data in this report whether statement samples smaller or larger than 100 would yield comparable results.

In the opinion of the writers several extensions of this research could be undertaken with profit:

a. The validity of the Climate Index needs to be pursued in terms of several criteria of teaching effectiveness, particularly pupil growth. In the absence of wholly adequate criteria it is important to determine some of the correlates of teachers' verbal behavior. Some of these factors could be subsumed under teacher personality and experiential background. Others might involve responses of classes and pupils and characteristics of the school situation.

b. Normative studies for determining the range of verbal behavior and classroom climate among types of teachers and among types of schools would be desirable. In this connection special attention must be paid to sampling procedures.

c. The consistency of the climate provided by a teacher may be an aspect of classroom behavior worthy of intensive study. It may be found that whether a teacher's average Climate Index is 25 or 75 is not as important as whether she tends to create a consistent atmosphere in which youngsters know what to expect. It can be hypothesized that consistency of classroom climate reduces anxieties on the part of pupils, and that such anxiety reduction facilitates social and intellectual growth.

d. Another extension, which should perhaps precede the others, involves the determination of the size of the most appropriate statement sample for picturing a teacher's verbal behavior. The arbitrary sample size of 100 statements employed in this study may account for some of the inconsistency in teachers' verbal behaviors from occasion to occasion. A study needs to be made in which varying numbers of statements are recorded and compared.

APPENDIX

Although the procedures involved in assessing classroom climate by categorizing the statements of the teacher have been described in the body of this report, the reader who wishes to utilize this method for research purposes will probably feel the need for a more thorough discussion. The material to be presented in this appendix is designed to satisfy this need, in part. For the most complete treatment, the reader is urged to consult articles by Withall (15, 16) and, in particular, his thesis (14).

A Suggested Procedure for Using Withall's Technique

There is no single procedure for developing skill in the categorization of verbal behavior which will prove best under all circumstances. The problem under investigation and the resources available to the investigators will, for the most part, suggest efficient procedures. It seems possible, however, to list certain principles which, under virtually all circumstances, research workers may adopt with profit. Much of the material to follow has been adapted from Withall's thesis, as modified in the light of the experience of the present authors.

Plan to work in teams, not alone. For many reasons it is wise in attempting to employ Withall's technique to work with others rather than alone. Only when categorizations have been made by at least two persons, working independently, is it possible to obtain a measure of agreement between judges. Moreover, in the early stages of practice with the technique, careful consideration of those statements on which disagreement is found will prove especially helpful in clarifying the criteria for the various categories. Needless to say, after considerable practice with the technique and satisfactory evidence of agreement with other judges, an individual judge may undertake categorizations alone.

Review the literature on the technique. A considered review of the literature will not only reveal the various problems to which Withall's technique has been applied, but will also indicate ways in which the data obtained by this technique may be treated statistically. An especially careful review should be made of the criteria for the seven statement categories. Each judge should attempt to develop a thorough understanding of what constitutes a statement and the defining properties of the various categories. Discussion among the judges at this stage will help prevent misconceptions and will tend to produce more uniform understandings of the category criteria.

Employ practice transcripts. Transcripts of class sessions should be obtained and used by judges as practice material. Withall's thesis contains a short practice transcript, and the two transcripts which follow may be used for practice purposes by masking the categorizations given each statement. All categorizations should be done independently so as to avoid contamination. Comparisons of the categorizations of different judges will tend to reveal significant areas of disagreement which will require restudy of category criteria. In the initial stages of skill development it will frequently be necessary to refer to the criteria of the seven categories as one actually categorizes a transcript. At later stages, however, this will prove less necessary and should be avoided.

Observe in schools. As soon as judges have developed some skill in categorizations, classroom observations in which verbatim records are made should be undertaken. As before, categorization should be based upon the transcribed record, but the categorizing process will present fewer difficulties since the judges were present during the actual class session. Discussions between observers will introduce additional clarifications.

Categorization may be attempted in the "live" situation. Under certain circumstances it may prove desirable to avoid any procedure for taking a verbatim record of what the classroom teacher says, and instead categorize her statements as she makes them, i.e., in the "live" situation. This should not be attempted unless the observers are thoroughly acquainted with the Withall technique, since it requires the ability to recognize and categorize statements with great rapidity. The advantage of this procedure lies partly in its simplicity, for it requires neither expensive, temperamental, sound-recording equipment nor a highly skilled stenographer. At the same time, by categorizing statements as they are made, the observer gains additional cues from the teacher's posture, gestures, facial expressions, and the like.

The presence of sound equipment or a recording stenographer in the classroom is often disturbing to the teacher and class. Under such circumstances a representative sample of a teacher's classroom behavior is difficult to obtain. In certain situations teachers or administrators will not permit verbatim records to be made. In other situations the presence of a stenographer or recording apparatus in the classroom will indicate to the teacher that her *verbal* behavior is of major concern to the investigators.

For these reasons the present writers recommend that Withall's technique be employed in the "live" situation, if the observers have sufficient skill to do so. The principal disadvantage of this method is that no permanent, accurate record of what was actually said by the teacher is available. This disadvantage will usually be minor, however, and may be partially overcome by having the observers write a brief description of the class session after it has transpired.

Transcriptions of Actual Class Sessions

Presented below are two transcriptions of portions of actual class sessions. The first, Transcript A, is taken with Professor Withall's permission from his thesis (14). Transcript B was obtained by the authors of this report in the early stages of their practice observations.

TRANSCRIPT A

(NOTE: Three diagonal lines at the end of a statement denote an interruption of the person speaking.)

Teacher C is giving individual help in a Grade VII mathematics class.

Category

1. 4 C: So many people are absent today that I question whether we should er work ahead on er anything else of ///
Pupil: Let's finish our graphs.
2. 5 C: I think the thing that you'd better do is, first, if you haven't finished these tables on heights and weights you'd better do those, both height and weight and the graph if you haven't finished that. If you have finished those, if you've finished the graphs, then, both the tables and the graphs, then ///
Pupil: Miss C.?
3. 4 C: Yes?
Pupil: Do we have to hand the weight in? I haven't finished height.
4. 3 C: No! I'm giving you time now to finish both tables, if you've finished both of those. . . .
Pupil: Miss C!
5. 4 C: Yes?
Pupil: (indistinguishable)
6. (a) 3 (b) 7 (c) 6 C: (a) Those are the graphs? I see. You were told to keep those until we can talk about them. (b) I intended to start the discussion of them today but so many are away that I think we hadn't better, uh, everybody, (c) Judy!! I've just said what? Huh?
Judy: I didn't hear . . . (indistinguishable)
7. 3 C: I said we could finish them right now. If everybody has finished both the tables and the graphs.
Pupil: Miss C, do we have to hand in both height and weight?
8. 3 C: Eventually.
Pupil: Not today, we don't have to hand them in today.
9. 6 C: No we're finishing them—you were absent, so you'll have to wait until I can explain to you what you are supposed to be doing.
Pupil: Shall I explain it to him?
10. (a) 1 (b) 5 C: (a) You can, if you can do it quietly. (b) Just a moment I think that if, er, if you've finished all this we've, we've done most of the work up to page 163 in the book, those who have finished all of this can start to study page 163 and start on that list of problems there. See?
Pupil: (indistinguishable)
11. (a) 3 (b) 1 C: (a) No I want them finished and then you're going to tell us about them orally. Just as soon as you've finished, maybe Friday, maybe not 'til Monday. See? But you'll get them done so that you'll have a story to tell. You see? (Pupil: Yah.) (b) All right. (No talking to be heard from anyone for fifteen to twenty seconds.)
12. 5 C: All right let's get these done now and if you've finished all of that start on page 163 in your book.
Pupil: Miss C. I'm putting that . . . (indistinguishable)
13. 1 C: All right.
Pupil: At first look it.
14. 3 C: What happened there?
Pupil: (indistinguishable)
15. 5 C: What, what happened there? Tell about it. What happened?
Pupil: It increased a lot.
16. (a) 3 (b) 7 C: (a) Yes, she, she got much heavier that year didn't she? (Pupil: Yes.) She gained a lot of weight. (b) I don't know why.
Pupil: Because you made it that way.
17. (a) 7 (b) 3 C: (a) No, I took these right out of the records. You want to see it it's in the book. I didn't make it that way that's the way it was. (b) Which girl you got?
Pupil: B
(Pupil comments are heard in far background.)
18. 3 C: This one, no. It must be right here. See it goes up.

Pupil: Hmmm. She must have increased that much.

19. (a) 6 C: (b) 5 (a) Now you see, this shows it better than your bar graph, (b) so while a bar graph's all right I would recommend that at least you fix the corners up as you suggested so that it looks like a line graph.

Pupil: (indistinguishable)

20. 3 C: Yes but this is, this is for the other one isn't it?

Pupil: Hummm.

21. (a) 7 C: (b) 5 (a) I'm saying that the bar graph makes it hard to see really what happened. (b) Huh? Why don't you make a line graph?

Transcript B

This transcript was prepared from stenographic notes of a teacher's verbalizations made during a typical class session. Statements made by pupils were not recorded by the stenographer. The explanatory comments noted within parentheses were taken from notes written by two observers during the same class session.

This sample of a teacher's verbal behavior was obtained during a short visit to a sixth-year classroom. It was the opinion of the observers that this teacher represented a "traditional" viewpoint in educational practice. There appeared to be considerable emphasis on recitation, deportment, and ready obedience to verbal instructions. The psychological intent of many of this teacher's statements may be unclear to the reader, but in the classroom situation intent was made apparent to the observers by means of accompanying gestures, inflections, and intonations. Many of the teacher's comments were sarcastic and "biting."

Approximately 25 pupils were present at the class session. Desks of the single and double variety were arranged in rows facing a blackboard which stretched across one end of the room. It was customary during this session for the teacher to walk back and forth in front of the rows of desks while conducting the class activities. The learning problem presented by the teacher during the observation concerned the development of the ability to tell time and to express time in writing. A large cardboard clock with adjustable hands hung from the blackboard and several similar clocks of the same construction were in the hands of the seated pupils. The teacher placed the hands of the clock to indicate "1:12" and had just asked for volunteers to write on the blackboard the time indicated by the clock.

Category

1. 5 "A!M!B! how about you?" (motioning toward the blackboard) (L. gingerly raises a hand)
2. 5 "L? All right, go ahead L., try it." (L., A., M., and B. all go to the blackboard)
3. 4 "Who else wants to try it?" (R. goes to the blackboard)
4. 6 (to L.) "We had 1:12, you wrote 1:00."
5. 5 "M! You try it . . . (M. begins to write on the board.) Let me see." (Teacher turns to R. who has been writing and erasing his work.)
6. 7 "R! *I* will give you just one more chance."
7. 5 "E! Go fix the clock again . . . the way you had it set before." (E. sets the clock to 11:30)
8. 3 "You see it's not twelve o'clock yet R."
9. 6 (annoyed) "Please put your hand down E."
10. 6 "All right, let's let A. have a chance and after that you pay a fine . . ." (A. writes on the board)
11. (a) 5 (a) "Leave that eleven . . . don't you
    (b) 1 move it A. (b) It's wonderful."
12. (a) 3 (to A.) "(a) How many minutes in an
    (b) 3 hour . . . (b) How many less 60 is it? . . . (A. responds correctly to both questions)
13. 1 (to A.) "Good girl! We congratulate you."
14. 1 (to M.) "You too! Good girl."
15. 4 "Now I am going to separate you." (Here the teacher was apparently referring to a frequently used device for dividing the class into two groups. The groups had titles paralleling those of well-known professional baseball teams.)
16. 5 "R! Would you get some chairs!"
17. 5 "D! You go get some chairs too!"
18. 4 "You may use these two tables . . ."
19. 5 "S! Pass the paper!"
20. 4 "Please don't take paper unless you're in the Giant group."
21. 5 (to D. who is carrying a chair) "Now you can bring it out." (The separation of the class was conducted with the speed and precision of military drill. One group remained in their seats while the other occupied chairs around tables at one side of the classroom. The basis for the division of the class was not apparent to the observers.)
22. 5 (to group at the tables working on arithmetic problems which were apparently being continued from the preceding day) "Now let me see, if I count three my Giants will be very busy . . . one . . ."
23. 5 (to group remaining in their seats)

"Please, Yankees put your books away . . ."

24. 5 "M! Move over there please."

25. 4 (absently to herself) "We need a few more chairs."

26. 5 (to Yankees) "All right now, these people move out this way . . ." (motioning) "One, two . . ."

27. 6 "E! in your own seat." (sharply)

28. 6 "Come on M., if you want to remain in this group, you have to be smart and not remain in your chair."

29. 4 "Now we are going to see how we do today."

30. 6 (to Giants, admonishing them) "All right, now you're going to be all by yourselves and you're going to be very smart."

31. 3 (to Yankees) "Now let me see if you can set the clock to the time we are supposed to be in school . . ."

32. 5 "M! You come and set the big clock."

33. 5 (helps M. adjust the hour hand) "Let's get this a little bit further down here . . ."

34. 6 (to M.) "You're five minutes late (for school)." (M. adjusts the clock)

35. 6 "Now you're too early!" (M. continues to manipulate the clock hands)

36. 1 "Now you're right." (R. takes over the hands of the clock)

37. 6 (to R.) "And you're late . . . now you're terribly late."

38. 6 (to R., sarcastically) "We will have to send you to the office."

39. 6 (to Giants who are whispering) "I'm afraid somebody will have to come out of that group."

REFERENCES

1. ANDERSON, H. H., & BREWER, J. E. *Studies of teachers' classroom personalities. App. Psychol. Monogr.*, 1946, No. 8. Stanford: Stanford Univer. Press.
2. BARTLETT, M. S. The use of transformations. *Biometrics*, 1947, **3**, 39-52.
3. COCHRAN, W. G. Some consequences when the assumptions for the analysis of variance are not satisfied. *Biometrics*, 1947, **3**, 22-38.
4. FISHER, R. A., & YATES, F. *Statistical tables for biological, agricultural and medical research.* (3rd Ed.) New York: Hafner, 1949.
5. FLANDERS, N. A. Personal-social anxiety as a factor in experimental learning situations. Unpublished doctoral dissertation, Univer. of Chicago, 1949.
6. KENDALL, M. G. *Rank correlation methods.* London: Griffin, 1948.
7. LINDQUIST, E. F. *Design and analysis of experiments in psychology and education.* Boston: Houghton Mifflin, 1953.
8. MITZEL, H. E. Interest factors predictive of teachers' rapport with pupils. Unpublished doctoral dissertation, Univer. of Minnesota, 1952.
9. ORLEANS, J. S., *et al.* Some preliminary thoughts on the criteria of teacher effectiveness. *J. educ. Res.*, 1952, **45**, 641-648.
10. PERKINS, H. The effects of social-emotional climate and curriculum on group learning of in-service teachers. Unpublished doctoral dissertation, Univer. of Chicago, 1949.
11. RABINOWITZ, W., & TRAVERS, R. M. W. Problems of defining and assessing teacher effectiveness. *Educ. Theory*, 1953, **3**, 212-219.
12. THELEN, H. A. Experimental research toward a theory of instruction. *J. educ. Res.*, 1951, **45**, 89-136.
13. THELEN, H. A., & WITHALL, J. G. Three frames of reference: the description of climate. *Hum. Relat.*, 1949, **2**, 159-176.
14. WITHALL, J. G. The development of a technique for the measurement of social-emotional climate in classrooms. Unpublished doctoral dissertation, Univer. of Chicago, 1948.
15. WITHALL, J. G. The development of a technique for the measurement of social-emotional climate in classrooms. *J. exp. Educ.*, 1949, **17**, 347-361.
16. WITHALL, J. G. Assessment of the social-emotional climates experienced by a group of seventh graders as they moved from class to class. *Educ. psychol. Measmt*, 1952, **12**, 440-451.

(Accepted for publication July 27, 1953)

Vol. 67, No. 19 Whole No. 369, 1953

Psychological Monographs: General and Applied

An Experimental Study of Extinction[1]

Marshall B. Jones

University of California at Los Angeles

Early in the study of intermittent reinforcement (3, 5, 6, 12) one conclusion emerged as a general law of learning: schedules of intermittent reinforcement result in greater resistance to extinction than do comparable schedules of continuous reinforcement.

Perhaps the most immediate consequence of this finding was that the then-current formulations of the law of effect had to be revised. Until the advent of intermittent reinforcement, the law of effect usually took something like the following form: habit strength and, therefore, its measure, resistance to extinction, are increasing functions of the number of reinforced responses. This supposition was untenable after the studies of intermittent reinforcement had become well known.

The theoretical attempts (1, 4, 9, 11) to rescue the law of effect were very nearly as numerous as the experimental studies which had placed the law in jeopardy. Nevertheless, a single explanation gradually became general: in intermittent reinforcement the acquisition conditions, since they involve a sizable number of extinction trials, are more similar to extinction conditions than are the acquisition conditions in continuous reinforcement. Hence, there is more stimulus generalization from acquisition to extinction in intermittent-reward training than in continuous-reward training. This greater stimulus generalization serves to perpetuate the intermittently acquired response after the continuously acquired response has extinguished. This theory, though derived from outright studies of intermittent reinforcement, has consequences which go far beyond its historical origins. The principal purpose of the present study is to test three of these more remote consequences.

Altogether there are in the present study four groups of experimental animals: escape, avoidance, intermittent escape, and limited avoidance. The extinction conditions for all four groups are the same; the acquisition conditions, however, differ. More specifically, the acquisition conditions for the three groups of nonescape animals are all more similar to extinction conditions than are the acquisition conditions for the escape animals. Whether the acquisition conditions for the three nonescape groups differ among themselves in terms of similarity we cannot say. All we can assert with confidence is that all three are more similar to extinction conditions than are the acquisition conditions for the escape animals. In conseqence, stimulus generalization from acquisition to extinction should be greater among the animals of the three nonescape groups. From this we may conclude that the avoidance, intermittent-escape, and limited-avoidance animals should all be more resistant to extinction than the escape animals.

---

[1] This monograph constitutes the essence of a doctoral dissertation done at the University of California at Los Angeles under the direction of Professor Howard C. Gilhousen, for whose encouragement and criticism the author is deeply grateful.

Of these three consequences, the first, that resistance to extinction should be greater among the avoidance than among the escape animals, has already been tested and confirmed by Sheffield and Temmer (10), who used a single, total-response measure. We, on the other hand, have broken up the total response into a starting and a running response. We are, therefore, in a position to ask the question: Does the starting or the running response undergo the greater deterioration in extinction? The answer to this question is given in the sections on extinction.

The theoretical issues underlying the sections on recovery and acquisition will be presented together with the results to which they relate.

I. APPARATUS

The apparatus employed in the present study was a grid runway leading to a wooden goal compartment. The grid measured 48 inches in length, 7 inches in width, and was surrounded on all sides, except for the escape "hole," by 18-inch-high walls. The escape hole, 6 inches in height, was cut in the wall at the far end of the grid. The hole led to the goal box, which was 1 foot deep, 7 inches wide, like the runway, and surrounded by 6-inch-high walls. The entire apparatus was made of plywood and was unpainted except for the goal box, which was painted black. A guillotine door at the entrance to the goal box prevented retracing.

One foot from the beginning of the grid two round holes were cut opposite each other in the side walls. These holes permitted a beam to an electric eye to pass over the runway. Just inside the goal box a second pair of holes permitted a second beam to pass over the runway. These two beams provided our response measures. When the animals were dropped on the grid, two electric clocks, calibrated in hundredths of a second, were activated. When the rat cut the first beam, the first clock stopped; when it cut the second beam, the second clock stopped. Hence, our two raw measures were starting time and total time.

The shock stimulator delivered 0.25 milliamperes to the grid. Since the resistance of the rat varies between 200 and 400 thousand ohms, the voltage drop across the shocked rat varied between 50 and 100 volts. These amperages and voltages are approximately standard in shock situations.

Connected with the shock stimulator was an intervalometer, which could be set so that at any desired delay after the clocks were activated shock would begin. In order to draw sufficient current through the relays that stopped the clocks, two light bulbs, used purely as resistors, had to be inserted in the circuit.

In Sheffield and Temmer's work the timing apparatus was silent. Ours was not. More specifically, two clicks were audible: the first sounded when the clocks were activated, and the second when the second beam was cut. The first click, since it coincided with dropping on the grid, was simply another cue to which the starting response could become conditioned. It, together with dropping and the grid compartment, constituted the stimulus complex that initiated each trial. The second click, since it coincided with entrance into the goal box, constituted, together with the goal box itself, the stimulus complex that ended the trial. While these clicks do not affect the theory of the experiment, they may very well have influenced the *absolute* levels of resistance to extinction observed. Sheffield and Temmer's rats showed less resistance to extinction than did ours. One plausible explanation for this difference is that the clicks of our apparatus provided a more stable stimulus configuration to serve, on the one hand, as the conditioned stimulus, and, on the other, as secondary reward.

II. SUBJECTS AND PROCEDURE

The subjects were 44 male, hooded rats. All the animals were born within a month of each other and all were approximately 200 days old at the start of the experiment. All had participated two months earlier in a thirst-motivated, bar-pressing experiment. None of the animals had had any previous experience of shock. The animals varied in weight from 270 to 417 grams. Prior to the experiment the animals were weighed and ranked in order of increasing weight. The animals were then divided into sets of four by weight, beginning with the four lightest animals and ending with the four heaviest. From each set one animal was assigned at random to each group. All four groups were, therefore, paired for weight.

Throughout the experiment the animals were maintained on an ad libitum schedule for both food and water. The experimental sessions were begun daily around noon. Two rats were run in each session. For the first 11 days one rat of the pair came from the escape group, and the second rat of the pair was his weight match from the limited-avoidance group. In the second 11 days one rat of the pair came from the avoidance group, while the second was its weight match from the intermittent-escape group.

There were certain procedural rules which applied equally to all groups. Each rat was placed on the grid every 2 minutes. This intertrial interval continued without disturbance from the beginning of acquisition until the animal was discontinued. Since the rats were run in pairs, one rat of the pair was placed on the grid every minute. All animals received 40 acquisition trials of their respective kinds, and then shock was withheld altogether. Each animal was run at least 10 trials in extinction. If the rat had not reached the criterion of three consecutive trials of 5 seconds or more in the first 10 trials of extinction, he was continued until he did or until he passed the maximum permitted. The maximum fell at 60 trials in extinction. If an animal was still running at the end of 60 extinction trials, he was discontinued. At no time, in acquisition or extinction, was an animal allowed to remain on the grid more than 10 seconds without disturbance. At 10 seconds he was gently paddled into the goal box. These maxima, 60 trials in extinction and 10 seconds in any one trial, had to be imposed for purely practical reasons. In each trial two measures were recorded: starting time, from dropping to the first beam; and running time, from the first to the second beam. Occasionally a rat would leap the first beam and its starting time for that trial would be lost. Whenever the occasional loss of starting time is of relevance to the presentation of results, the treatment of the lost trials will be discussed.

The escape animals in acquisition always landed on a charged grid. Since the grid remained charged, their only escape was through the escape hole into the goal box. At the 41st trial, shock was turned off and the trials continued as before until the rat was discontinued by criterion or by maximum.

The avoidance animals landed on an uncharged grid. Two seconds later, if they had not reached the goal box, shock set in. Throughout the 40-trial acquisition period the intervalometer was set at 2 seconds. At the 41st trial shock was permanently withheld and the animal was run until discontinued by criterion or by maximum. One thing is quite clear: the acquisition conditions for the avoidance animal were considerably more similar to extinction conditions than were the acquisition conditions for the escape animal. In the first place, the avoidance animal on many an acquisition trial avoided shock altogether. When it did, the trial was, for all intents and purposes, identical to an extinction trial, since no shock was experienced. In the second place, when the avoidance animal did receive shock, it received it from an initially uncharged grid. For the avoidance animal the grid was always uncharged initially. In contrast, the escape animal never experienced in acquisition a trial free from shock or an initially uncharged grid.

The intermittent-escape animals, when they received shock, received it in the same way as did the escape animals, i.e., from the moment of landing. On their nonshock trials the grid was initially and permanently uncharged. As acquisition progressed, the shock trials became increasingly less frequent. Shock was presented on trials 1-5, 7, 10, 14, 19, 25, 32, and 40. At the 41st trial shock was turned off and the animals were run until discontinued by criterion or by maximum. As we have seen, the greater similarity of avoidance training consists essentially of two elements: acquisition trials entirely free from shock, and receipt of shock, when experienced at all, from an initially uncharged grid. Clearly, the second element of similarity in avoidance training, receipt of shock from an initially uncharged grid, is altogether absent. The first element, however, interspersal in acquisition of shock-free trials, is decidedly present.

For the limited-avoidance animal shock began 0.8 second after his landing on the grid. This interval is too short for the rat to avoid shock altogether; it simply can't cross the grid that fast. For this reason the first element of similarity in avoidance training, interspersal in acquisition of shock-free trials, is entirely excluded. The second element, however, is present. The limited-avoidance animal always received shock from an initially uncharged grid. The limited-avoidance animals, like the rest, received 40 acquisition trials, after which shock was turned off and the animals were run until discontinued by criterion or by maximum.

III. STATISTICAL ANALYSIS

In treating the experimental results two techniques are employed: the Mann-Whitney *U* test (8) and the Wilcoxon test for paired scores (13). Both tests are nonparametric ranking techniques. Furthermore, both are designed to test the null hypothesis that there is "no mean difference between the two groups involved."

In evaluating our results we have used one-tailed tests exclusively. In justification of this usage we may say that we have always had theoretical and empirical reason to expect the difference to lie in the observed direction.

Although the Mann-Whitney test is designed for unpaired scores, it may be applied to paired data. In doing so, one arbitrarily fails to remove a source of variance which may result in nonsignificance when significance actually obtains. Whether or not one loses appreciable significance in applying the Mann-Whitney test to paired data depends upon the correlation, positive or negative, between the paired scores. If there is very little or no correlation, the loss is negligible. There are in the present study two types of pairing, by weight and by animal identity. In some instances the Mann-Whitney test had to be applied to data which were initially paired for weight, because one or more animals of one of the groups had to be discarded. We have, therefore, for the sake of uniformity, treated all data paired only by weight with the Mann-Whitney test. Comparison, when possible, with the Wilcoxon test shows negligible loss of significance. However, with scores paired on the basis of the animal which produced them, the Wilcoxon test has to be applied. Otherwise, the loss of significance is ruinous.

When measures are used that have maximum values, for example, 60 trials in extinction, all animals which passed the maximum were given the same score and rank. We would, of course, have preferred to do without maxima. However, granted their practical necessity, this convention seems to be the most reasonable treatment of the data.

Since only two tests of significance are used, we do not specify the test used when quoting a significance level. However, a glance at the appropriate table will quickly settle the matter.

IV. EXTINCTION

Escape vs. Avoidance Training

Avoidance training presents in acquisition a set of conditions which are considerably more similar to those which obtain when extinction is introduced than does escape training. The higher degree of similarity consists essentially of two elements: (*a*) the interspersal in acquisition of trials entirely free from shock, and (*b*) receipt of shock, when experienced at all, from an initially uncharged grid. The presence of these two elements in avoidance training and their absence in escape training should, through stimulus generalization, produce greater resistance to extinction among the avoidance animals than among the escape animals.

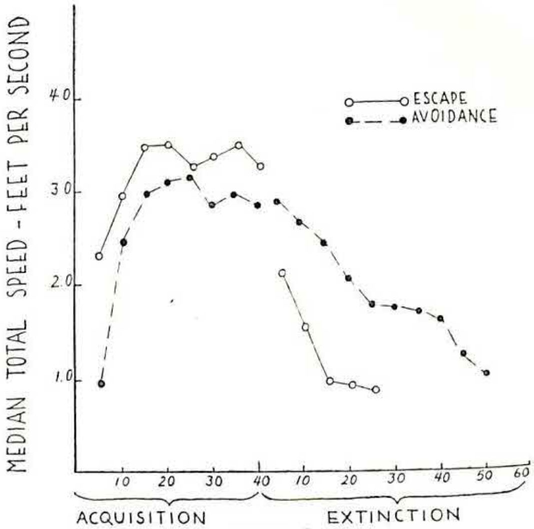


FIG. 1. Median Total Speed for Escape and Avoidance Training in Acquisition and Extinction.

In Figure 1 the median total speeds for the two groups are plotted. To obtain these curves we *(a)* calculated the median total speed in each consecutive set of five trials for each animal, and *(b)* calculated the medians for each group from the individual medians for each 5-trial set. The group medians appear in Figure 1 plotted at the last trial of the 5-trial set in question. When, in extinction, an animal had been discontinued, his total speed was arbitrarily supposed to lie below the group median for all 5-trial sets to follow. When the median animal, i.e., the sixth, extinguished, the group was supposed to extinguish and no point was plotted.

As is apparent from Figure 1, the extinction curve of the avoidance animals is considerably longer than that of the escape animals. Table 1 presents the statistical evaluation of this difference in terms of trials to the criterion of three consecutive trials of 5 seconds or more. The result shows that the difference is significant beyond the .01 level. We may, therefore, conclude that the similarity theory of extinction finds confirmation in the comparison of escape with avoidance training.

TABLE 1

ESCAPE VS. AVOIDANCE TRAINING: TRIALS TO THE CRITERION OF THREE CONSECUTIVE 5-SECOND TRIALS

Trials	Rank	Group
4	1	E
6	2	E
11	3	E
14	4	E
16	5	E
26	6.5	E
26	6.5	A
29	8	A
33	9.5	E
33	9.5	A
35	11	E
37	12	E
40	13	A
48	14	A
53	15	A
55	16	E
59	17	A
60+	20	E
60+	20	A
60+	20	A
60+	20	A
60+	20	A

Σ (A ranks) $= 163.0$, $U = 24.0$, $\bar{U} = 60.5$, $\sigma_u = 15.65$, $CR = 2.33$, $p < .01$.

Does the starting or the running response undergo the greater deterioration in extinction? To answer this question the median starting times in the first five and in the last five trials of extinction were calculated for each animal. When an animal ran past the maximum, his "last five trials" were trials 56–60. The group medians were then calculated from the individual medians. In the same way the median running times in the first five and in the last five trials of extinction were calculated for each group. Among the escape animals the median starting time in the first five trials was 0.31 second and in the last five trials 0.35 second. This increase is very slight indeed when compared to the rise in running time, which increased from 1.54 seconds in the first five trials to 4.86 seconds in the last five. Statistics are hardly necessary: among the escape animals the extinction effect was felt, almost entirely, in the running response. Even when the escape animal stopped altogether further down the grid, he still started, and with very nearly as much alacrity as he did at the beginning of extinction. The avoidance animals present a very similar picture. Their starting times increased from 0.25 second to 0.38 second while their running times increased from 1.11 seconds to 3.08 seconds. Once again, statistical analysis is unnecessary. Among the avoidance animals, as among the escape animals, the running response was almost exclusively responsible for the total extinction effect.

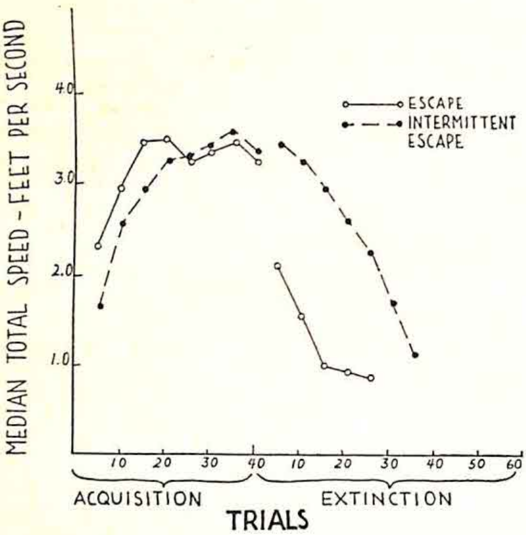


FIG. 2. Median Total Speed for Escape and Intermittent-Escape Training in Acquisition and Extinction.

Escape vs. Intermittent-Escape Training

The interspersal, in acquisition, of trials entirely free from shock is the first element of similarity in avoidance training. An intermittent-escape schedule involves this element to the exclusion of the second, receipt of shock from an initially uncharged grid. The present comparison, therefore, enables us to examine in isolation the extinction effects of the first element of similarity in avoidance training. These effects, according to the similarity theory, should include a greater resistance to extinction among the intermittent-escape animals than among the escape animals.

In Figure 2 the median total speeds for the two groups are presented. These curves were plotted in the same way as were the curves in Figure 1. Since the intermittent-escape animals received their escape training on trials 1–5, 7, 10, 14, 19, 25, 32, and 40, grouping the trials by 5's destroys one of the principal characteristics of their acquisition curve: a scalloping peaked on the shock trials. Since, however, our present concern is with extinction, we will defer a detailed treatment of this phenomenon to the section on acquisition. The intermittent-escape group originally contained 11 animals. Of the 11 animals, 8, those whose curve is plotted in Figure 2, displayed an increasingly stable starting-and-running response throughout acquisition. The remaining 3 animals extinguished in acquisition. This being the case, the comments on extinction which follow apply only to animals which responded in acquisition in one of two quite disparate ways. We will, when we take up acquisition, discuss these two patterns in greater detail.

Figure 2 shows a longer curve in extinction for the intermittent-escape animals. This difference in terms of trials to the criterion of three consecutive trials of 5 seconds or more is significant at the .025 level (Table 2). We may, therefore, safely say that the comparison of escape and intermittent-escape training lends additional support to the similarity theory of extinction.

We have already seen that the escape

TABLE 2

ESCAPE VS. INTERMITTENT-ESCAPE TRAINING: TRIALS TO THE CRITERION OF THREE CONSECUTIVE 5-SECOND TRIALS

Trials	Rank	Group
4	1	E
6	2	E
11	3	E
14	4	E
16	5	E
26	6	E
28	7	I
29	8	I
33	9	E
34	10	I
35	11	E
37	12	E
38	13	I
53	14	I
55	15	E
56	16	I
60+	18	E
60+	18	I
60+	18	I

Σ (I ranks) $= 104.9$, $U = 20.0$, $\bar{U} = 44.0$, $\sigma_u = 12.11$, $CR = 1.98$, $p < .025$.

and avoidance animals extinguish almost exclusively in the running response. In this respect, intermittent-escape training is no different. Their starting times increased from 0.22 second to 0.26 second from the first five to the last five trials of extinction. Meanwhile, their running times rose from 0.92 second to 5.90 seconds. Whether we consider absolute or relative increase, the starting response is by far more resistant to extinction than the running response.

Escape vs. Limited-Avoidance Training

For the limited-avoidance animals, shock set in 0.8 second after they hit the grid. Since under this condition the rats could not entirely avoid shock, and since they received shock from an initially uncharged grid, the limited-avoidance group represents the second element of similarity in avoidance training, uncontaminated by the first. A theory which implies that resistance to extinction increases with the similarity between acquisition and extinction conditions clearly requires the limited-avoidance animals to run longer in extinction than the escape animals.

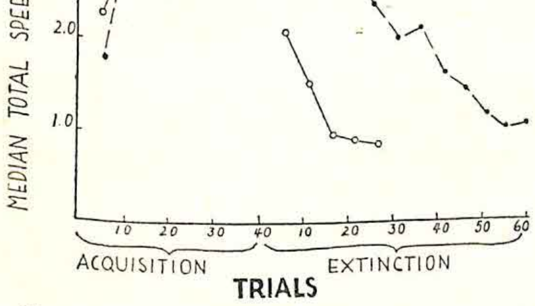


FIG. 3. Median Total Speed for Escape and Limited-Avoidance Training in Acquisition and Extinction.

TABLE 3

ESCAPE VS. LIMITED-AVOIDANCE TRAINING: TRIALS TO THE CRITERION OF THREE CONSECUTIVE 5-SECOND TRIALS

Trials	Rank	Group
4	1	E
6	2	E
11	3	E
14	4	E
16	5.5	E
16	5.5	L
20	7	L
26	8	E
32	9	L
33	10	E
35	11	E
37	12	E
47	13	L
55	14	E
60	18.5	E
60	18.5	L
60	18.5	L
60	18.5	L
60	18.5	L
60	18.5	L
60	18.5	L
60	18.5	L

Σ (L ranks) $=164.0$, $U=23.0$, $\bar{U}=60.5$, $\sigma_u=15.65$, $CR=2.40$, $p<.01$.

In Figure 3 the acquisition and extinction curves are plotted in the same way as in the first two figures.

In terms of trials to the criterion of three consecutive trials of 5 seconds or more, the similarity theory of extinction is confirmed (Table 3). The limited-avoidance animals ran longer in extinction than did the escape animals. Significance reaches the .01 level.

All three groups examined so far have shown the effects of extinction almost entirely in the running response. The limited-avoidance animals complete the picture. Their median starting times increased from 0.31 second to 0.42 second over the extinction period. Meanwhile, their median running times increased from 1.16 seconds to 3.74 seconds. In all four groups, then, the starting response was decidedly more resistant to extinction than was the running response.

V. RECOVERY

Escape and Intermittent-Escape Training

a. The starting response. Estes in his experimental study of punishment (2) presented a wide variety of results in support of the hypothesis that a response which has been continuously or intermittently punished will, if performed and not punished, recover in strength. At first blush, this finding might seem of little relevance to the present experiment. On closer analysis, however, recovery is seen to play a crucial role in the behavioral systems under investigation.

Since the words "punishment" and "punished" figure prominently in our analysis, a brief account of the way we use them may be in order. The word "punishment" applies to *stimuli,* like shock, which, quite apart from learning, produce in an animal the usual signs of fear: defecation, urination, quivering, squealing, etc. The word "punished," on the other hand, applies to *responses.* Whether or not a response is punished depends upon the *difference* in punishment before and after the response. The simple fact that punishment follows a response does not mean that the response is punished. A response is punished only if its performance results in "a turn for the worse."

With this terminological understanding we turn to the starting response among the escape and intermittent-escape animals. More specifically, we are concerned with the shock trials in the schedules of these two groups. Typically, the animals of all four groups landed on the grid with their hind legs and leaped forward. If the two hind legs fell on even-spaced spokes, they drew no current and the animal did not receive shock until after his initial thrust had already carried him beyond the first beam. If the two hind legs fell on odd-spaced spokes, they drew current, but the shock experienced was surely less than that which obtained after the starting response had been completed and all four legs were in contact with the grid. In short, the starting response among the escape and intermittent-escape animals was punished in every shock trial. On the nonshock trials punishment neither preceded nor followed the starting response. Hence, the starting response among the escape and intermittent-escape animals went unpunished in their nonshock trials.

Since for the escape animals, every acquisition trial was a shock trial, the starting response among the escape animals was punished in every trial up to and including the 40th. From the 41st trial on, the starting response was no longer punished. Hence, if Estes' conclusions hold generally, the starting response among the escape animals should, beginning with the 41st trial, gradually recover in strength, i.e., the corresponding starting times should gradually decrease.

The starting response among the intermittent-escape animals was, as we have seen, punished in trials 1–5, 7, 10, 14, 19, 25, 32, and 40, and unpunished in trials 6, 8–9, 11–13, 15–18, 20–24, 26–31, and 33–39. Let us call these latter series of 1, 2, 3, 4, 5, 6, and 7 trials "null" series. Clearly, each null series follows a series of trials in which the starting response has been continuously or intermittently punished. We should, therefore, expect to find recovery of the starting response in the null series of the acquisition schedule. More concretely, starting times should tend to decrease during the null series of intermittent-escape training.

If recovery of the starting response is manifest throughout acquisition under intermittent-escape training but delayed under escape training until the 41st trial, then the starting response among the intermittent-escape animals should, toward the end of acquisition, be stronger, i.e., starting times should be lower than they are among the escape animals.

To test the first consequence of the foregoing analysis we compared the median starting time (Mdn_1) in the first five trials of extinction with the median starting time (Mdn_2) in the second five trials of extinction for each escape animal. If recovery is indeed present, the medians Mdn_2 should be less than the medians Mdn_1. This is, in fact, the case. Significance reaches beyond the .01 level (Table 4).

To test the second consequence, we calculated for each intermittent-escape animal the median starting time in the first and in the second halves of each null series, with the obvious exception of the first. From these we calculated, again for each animal, the median Mdn_1 of the first half medians and the median Mdn_2 of the second half medians. If recovery occurs in the null series of the intermittent-escape schedule, the Mdn_2 medians should be less than the Mdn_1 medians. The results bear out the analysis. Significance reaches beyond the .05 level (Table 5). In 6 out of 48 null series at least one starting time was lost; the rat jumped the beam. Since the measure employed is very delicate, these series were thrown out in preference to estimating the medians from the trials that remained.

TABLE 4

ESCAPE TRAINING: MEDIAN STARTING TIMES IN THE FIRST FIVE VS. THE SECOND FIVE TRIALS OF EXTINCTION

Pair	1st Five Trials	2nd Five Trials	Diff.	Rank
1	.61	.65	−.04	3.5
2	.41	.31	+.07	6
3	.33	.21	+.08	7.5
4	.27	.22	+.05	5
5	.29	.25	+.04	3.5
6	3.60	3.60	.00	
7	.29	.30	−.01	1
8	.31	.23	+.08	7.5
9	.28	.25	+.03	2
10	.57	.48	+.09	9
11	.31	.19	+.12	10

Σ (negative ranks) = 4.5, $p < .01$.

TABLE 5

INTERMITTENT-ESCAPE TRAINING: MEDIAN STARTING TIMES IN THE FIRST HALF VS. THE SECOND HALF OF THE NULL SERIES

Pair	First Half	Second Half	Sign	Rank
1	.34	.27	+	8
2	.23	.22	+	3
3	.26	.25	+	3
4	.28	.24	+	7
5	.25	.24	+	3
6	.27	.25	+	6
7	.29	.30	−	3
8	.25	.24	+	3

Σ (negative ranks) = 3.0, $p < .05$.

To test the third consequence of our analysis, we compared the median starting times of the escape animals with those of the intermittent-escape animals in the last five trials of acquisition. The intermittent-escape animals definitely started faster. The result is significant at the .01 level (Table 6).

In Figure 4 median starting speeds in acquisition and the first ten trials of extinction are plotted for the escape and intermittent-escape groups. The curves are plotted in sets of five trials as were the earlier curves for total speed. In Figure 4 the first and third consequences of our analysis may be seen graphically.

In explaining Estes' results we have argued in another place (7) that an ani-

TABLE 6

Escape vs. Intermittent-Escape Training: Median Starting Times in the Last Five Trials of Acquisition

Seconds	Rank	Group
0.21	1.5	I
0.21	1.5	I
0.22	3	I
0.23	4.5	I
0.23	4.5	I
0.24	6	I
0.26	7	E
0.27	9.5	I
0.27	9.5	E
0.27	9.5	E
0.27	9.5	E
0.28	12	E
0.30	13	I
0.32	14	E
0.33	15	E
0.34	16	E
0.40	17.5	E
0.40	17.5	E
0.61	19	E

Σ (E ranks) $= 146.5$, $U = 7.5$, $\bar{U} = 44.0$, $\sigma_u = 12.11$, $CR = 3.01$, $p < .01$.

mal which has been punished, continuously or intermittently, for performing a given response in a particular situation acquires in consequence an *expectation* that the response in question will be punished when performed in that situation. If, having acquired such an expectation, the animal performs the response and is not punished, he is *effectively* rewarded. As a result, the response recovers in strength. This view of recovery is clearly adequate to the present results as well as to Estes' original findings. The escape animals, for example, build up in acquisition an expectation that they will be punished for starting. When, in extinction, they start and are not punished, their starting response is effectively rewarded. In consequence, their starting response recovers in strength, their starting times gradually decrease. An altogether similar argument applies to recovery in the null series of the intermittent-escape schedule. The third consequence of our analysis follows, as we have seen, from the first two and, hence, from the interpretation of recovery as the result of effective reward.

b. The running response. The escape and intermittent-escape animals experienced two sorts of trials: shock trials in which the animals landed on a permanently charged grid, and nonshock trials in which the grid was never charged. In the first case shock preceded but did not follow the running response. In the second case shock neither preceded nor followed the running response. In neither case, then, was the running response punished. Hence, the running response could hardly show recovery and, as a matter of fact, never did.

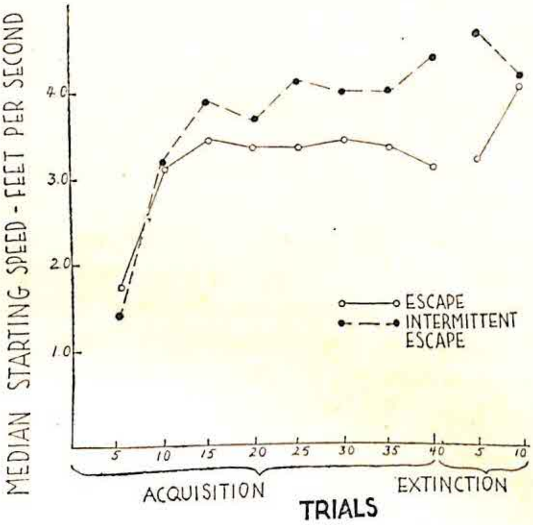


Fig. 4. Median Starting Speed for Escape and Intermittent-Escape Training in Acquisition and the First Ten Trials of Extinction.

The presence of recovery in the starting response and its total absence in the running response provide, in our opinion, an explanation of the earlier finding that the starting response in these two groups was decidedly more resistant to extinction than was the running response. In "extinction" the starting response actually increased its strength, at least for awhile. In contrast, the running response, unbolstered by recovery, deteriorated from the outset of the extinction schedule.

Avoidance and Limited-Avoidance Training

a. The starting response. The starting response among the limited-avoidance animals, just as among the escape animals, was punished in every acquisition trial. In this respect the sole difference between the two groups was that the starting response received shock immediately among the escape animals but only after a delay of approximately 0.5 second among the limited-avoidance animals. Among the avoidance animals the starting response was also punished at a delay. However, in many, if not most, of their acquisition trials the avoidance animals avoided shock altogether, and the starting response went unpunished. In neither group was the starting response punished under the extinction schedule. In brief, then, the two avoidance groups were, with respect to recovery, altogether analogous to the two escape groups, the delay factor, of course, excepted.

By virtue of the analogy, three parallel consequences follow: (*a*) the starting response among the limited-avoidance animals should show recovery in the first ten trials of extinction, (*b*) the starting response among the avoidance animals should show recovery during the null series of their acquisition schedule, i.e., during those trials in which they avoided shock altogether, and (*c*) the starting response among the avoidance animals should, toward the end of acquisition, be stronger than among the limited-avoidance animals.

TABLE 7

LIMITED-AVOIDANCE TRAINING: MEDIAN STARTING TIMES IN THE FIRST FIVE VS. THE SECOND FIVE TRIALS OF EXTINCTION

Pair	1st Five Trials	2nd Five Trials	Diff.	Rank
1	.50	.43	+.07	8
2	.20	.20	.00	
3	.47	.44	+.03	3.5
4	.22	.23	−.01	1.5
5	.30	.34	−.04	5
6	.28	.31	−.03	3.5
7	.32	.26	+.06	6
8	.26	.25	+.01	1.5
9	.31	.24	+.07	7
10	.53	.35	+.18	9
11	.65	.21	+.44	10

Σ (negative ranks) $= 10.0$, $p < .05$.

TABLE 8

AVOIDANCE VS. LIMITED-AVOIDANCE TRAINING: MEDIAN STARTING TIMES IN THE LAST FIVE TRIALS OF ACQUISITION

Seconds	Rank	Group	Seconds	Rank	Group
.22	2	A	.30	11.5	L
.22	2	A	.32	13	L
.22	2	A	.33	14.5	A
.24	4	A	.33	14.5	L
.26	5	A	.35	16	L
.27	7	A	.37	17	L
.27	7	L	.42	18	A
.27	7	L	.44	19	L
.28	9	L	.45	20.5	L
.29	10	A	.45	20.5	L
.30	11.5	A	.48	22	A

Σ (L ranks) $= 155.0$, $U = 32.0$, $\bar{U} = 60.5$, $\sigma_u = 15.65$, $CR = 1.82$, $p < .04$.

To test the first consequence we compared median starting times among the limited-avoidance animals in the first five and in the second five trials of extinction. The result confirms the presence of recovery at the .05 level (Table 7).

We have made every effort to find an adequate measure of recovery in the acquisition series of the avoidance animals. However, the extreme irregularity in shock pattern from animal to animal frustrated our every attempt. A direct approach to recovery in acquisition among the avoidance animals has, therefore, been abandoned.

We may, however, approach the problem indirectly. The third consequence of our analysis follows, as we have seen, from the first two. Since the first conseqence was independently confirmed, direct confirmation of the third constitutes at the same time indirect support for the second. Comparing median starting times in the last five trials of acquisition, we found that the avoidance animals did indeed start faster than the limited-avoidance animals. Significance reaches beyond the .04 level (Table 8). This result confirms the third consequence and, hence, the second indirectly.

In Figure 5 median starting speeds in acquisition and the first ten trials of extinction are plotted for the avoidance and limited-avoidance animals. The role of recovery may be clearly seen.

The results of the last section fell quite naturally under the rubric of recovery as the result of effective reward. The results of this section may be analyzed in almost the same words. The starting response among the limited-avoidance animals was punished at a delay throughout acquisition. When, in extinction, the limited-avoidance animal starts and is not punished, starting is effectively rewarded. In consequence, the starting response recovers in strength. A null series of avoidance training, i.e., a series of trials in which the animal successfully avoids shock, always follows a series of trials in which the starting response has been continuously or intermittently punished at a delay. When the avoidance animal starts and is not punished (because he successfully avoids), the starting response is effectively rewarded. Recovery follows as a result. The third consequence follows from the first two and, hence, from our view of recovery, as the result of effective reward.

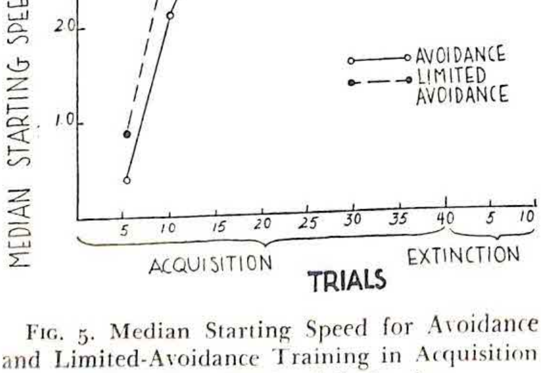


Fig. 5. Median Starting Speed for Avoidance and Limited-Avoidance Training in Acquisition and the First Ten Trials of Extinction.

b. The running response. In the last section we saw that neither escape schedule *allowed* the running response to be punished. In the two avoidance groups, on the other hand, the possibility exists that the running response may be followed and not preceded by shock. The limited-avoidance animals, for example, might receive the onset of shock while in a full run toward the goal box. Were this to take place with any consistency, we would expect to find evidence of recovery in the running response when shock was removed. However, though the possibility exists, the running response was not consistently punished in either avoidance group.

Each individual limited-avoidance animal tended in acquisition to a fairly constant total starting-and-running time. In consequence, he received the onset of shock from trial to trial at approximately the same point along the grid. The typical result was a pause just before the point where shock usually set in. This tendency, as we shall see in the next section, was sufficiently powerful to produce a significant increase in running time for the limited-avoidance animals toward the end of acquisition. In any case, the running response, at least in the later stages of acquisition, was not consistently punished. Pausing, however, was. In consequence, removal of shock, so far from producing recovery in the running response, should have produced recovery in the pausing response, i.e., confirmed the animals in their tendency to pause toward the end of the grid. What has been said of the limited-avoidance animals applies to the avoidance animals as well. They too slowed-up or stopped when shock set in, particularly in the later stages of acquisition. In consequence, slowing-up, not running, was confirmed by recovery when shock was omitted. Hence, as we should have expected, recovery of the running response was altogether absent among the animals of the two avoidance groups.

These observations enable us to understand the fact, previously reported, that the starting response among the animals of the two avoidance groups was distinctly more resistant to extinction than was the running response. The starting response in these two groups, just as in the two escape groups, was preserved and strengthened by recovery, while the running response was not. In consequence, the running response was less able to resist extinction and deteriorated first.

VI. ACQUISITION

Avoidance and Limited-Avoidance Training

Sheffield and Temmer reported that their avoidance animals showed a tendency to slow up in the later stages of acquisition. A glance at Figure 1 reveals that our animals exhibited a similar tendency. The median total time increased from 1.31 seconds in trials 16–20 to 1.43 seconds in trials 36–40. In terms of components, median running times increased from 1.02 to 1.18 seconds, while median starting times *decreased* from 0.30 to 0.27 second for the same sets of trials. The slowing-up was, therefore, a phenomenon confined to the running response. Table 9 reveals the increase in running time as significant at the .05 level. To understand the result we need only note that the avoidance animal typically started avoiding around the 10th trial. For some time afterwards, his running times remained very low. Then they started to increase, as a simple matter of extinction, until toward the end of acquisition he picked up a few more shocks. The animals varied, of course, as to when they started avoiding and how long they continued to avoid once they had started. (On the average, however, their running times were lowest in trials 16–20 and highest, from then on in acquisition, in trials 36–40.) This pattern of shock and avoidance accounts in the main for the result reported above.

Largely as a consequence of the first result, Sheffield and Temmer reported that the escape animals were running faster toward the end of acquisition than

TABLE 9

AVOIDANCE TRAINING: MEDIAN RUNNING TIMES IN TRIALS 16–20 VS. TRIALS 36–40

Pair	Trials 16–20	Trials 36–40	Diff.	Rank
1	.82	.72	−.10	3.5
2	.75	1.02	+.27	8
3	1.21	1.22	+.01	1
4	.82	.92	+.10	3.5
5	1.56	1.27	−.29	9
6	1.00	1.18	+.18	7
7	.96	1.10	+.14	5.5
8	.82	1.20	+.38	10
9	1.26	2.24	+.98	11
10	1.02	1.06	+.04	2
11	1.06	1.20	+.14	5.5

Σ (negative ranks) = 12.5, $p < .05$.

VII. CONCLUSION

Let us summarize the theory and results which have been our concern in the preceding pages. Out of the conflict between intermittent reinforcement and early formulations of the law of effect emerged a theory of extinction. This theory, reduced to testable essentials, is that *resistance to extinction increases with the similarity between acquisition and extinction conditions.* While this hypothesis had been widely checked in outright studies of intermittent reinforcement, its consequences for escape and avoidance schedules had gone unexamined. More specifically, the acquisition conditions of avoidance, intermittent-escape, and limited-avoidance training are all more similar to extinction conditions than are the acquisition conditions of escape training. Hence, the similarity theory of extinction entails three definite consequences: avoidance, intermittent-escape, and limited-avoidance training should all prove to be more resistant to extinction than escape training. All three consequences were confirmed by the experimental results.

The total response of reaching the goal box is analyzable into two components, a starting and a running response. In all four groups the extinction effect was felt almost exclusively in the running response. The starting response remained virtually intact long after the running response had almost entirely disappeared. An explanation of this finding was delayed until after the analysis of recovery had been presented.

An analysis of the starting response among the escape and intermittent-escape animals led to three definite expectations, all of which were supported by the evidence:

1. The starting response among the escape animals should show recovery in the early stages of extinction.

2. The starting response among the intermittent-escape animals should show recovery during the null series of their acquisition schedule.

3. The intermittent-escape animals should, in the later stages of acquisition, start faster than the escape animals.

Since the running response among the escape and intermittent-escape animals was never punished, there could be no recovery of the running response after punishment. The facts of recovery in the starting response and no recovery in the running response were offered as an explanation of the earlier result that the starting response was far more resistant to extinction than the running response in these two groups of animals.

The starting response among the avoidance and limited-avoidance animals was then subjected to an analysis altogether analogous to that which had been applied earlier to the starting response among the escape and intermittent-escape animals. The analysis issued in three analogous expectations:

1. The starting response among the limited-avoidance animals should show recovery in the early stages of extinction.

2. The starting response among the avoidance animals should show evidence of recovery during the null series of their acquisition schedule.

3. The starting response among the avoidance animals should be, toward the end of acquisition, faster than among the limited-avoidance animals.

Because of the irregularity of shock trials among the avoidance animals no adequate measure of recovery during their acquisition series could be found. As a result, consequence 2 could not be tested directly. The remaining two consequences were, however, confirmed.

The running response among the avoidance and limited-avoidance animals was rarely punished. In general, the animals were slowing up or stopping when shock set in, particularly in the later stages of acquisition. In consequence, recovery, so far from strengthening the running response, strengthened the slowing-up and stopping responses.

The absence of recovery in the running response and the presence of recovery in the starting response among the avoidance and limited-avoidance animals were offered in explanation of the decidedly greater resistance to extinction of the starting response among the animals of these two groups.

The results involving recovery were seen to follow from the interpretation of recovery as the result of effective reward.

The running response of the avoidance animals tended to slow up toward the end of acquisition. Largely in consequence of this result, the escape animals were running faster than the avoidance animals in the later stages of acquisition. The tendency of the limited-avoidance animals to pause just before their usual shock point produced a significant slowing-up in their running response toward the end of acquisition.

The group of three intermittent-escape animals which "extinguished in acquisition" was seen to have been a group in which conditioning of the starting-and-running response was faulty in the first five trials of acquisition. Finally, the dominant group of eight displayed two distinct acquisition curves: one for the shock trials and one for the nonshock trials.

REFERENCES

1. Denny, M. R. The role of secondary reinforcement in a partial reinforcement learning situation. *J. exp. Psychol.*, 1946, **36**, 373-389.
2. Estes, W. K. An experimental study of punishment. *Psychol. Monogr.*, 1944, **57**, No. 3 (Whole No. 263).
3. Grosslight, J. H., & Child, I. L. Persistence as a function of previous experience of failure following success. *Amer. J. Psychol.*, 1947, **60**, 378-387.
4. Hull, C. L. *Psychology seminar memoranda, 1940-1941.* Bound mimeographed copies on file in the libraries of the Univer. of Iowa, Univer. of North Carolina, and Yale Univer.
5. Humphreys, L. G. The effect of random alteration of reinforcement on the acquisition and extinction of conditioned eyelid reactions. *J. exp. Psychol.*, 1939, **25**, 141-158.
6. Jenkins, W. O., McFann, H., & Clayton, Frances L. A methodological study of extinction following a periodic and continuous reinforcement. *J. comp. physiol. Psychol.*, 1951, **44**, 155-167.
7. Jones, M. B. Effect, change, and expectation of reward. *Psychol. Rev.*, 1952, **59**, 227-233.
8. Mann, H. B., & Whitney, D. R. On a test of whether one of two random variables is statistically larger than the other. *Ann. math. Statist.*, 1947, **18**, 50-60.
9. Miller, N. E. Learnable drives and rewards. In S. S. Stevens (Ed.), *Handbook of experimental psychology*. New York: McGraw-Hill, 1951. Pp. 435-473.
10. Sheffield, F. D., & Temmer, Helena W. Relative resistance to extinction of escape training and avoidance training. *J. exp. Psychol.*, 1950, **40**, 287-298.
11. Sheffield, Virginia F. Extinction as a function of partial reinforcement and distribution of practice. *J. exp. Psychol.*, 1949, **39**, 511-525.
12. Skinner, B. F. *The behavior of organisms.* New York: Appleton-Century, 1938.
13. Wilcoxon, F. Individual comparisons by ranking methods. *Biometrics Bull.*, 1945, **1**, 80-83.

(Accepted for publication July 20, 1953)

Vol. 67, No. 20 Whole No. 370, 1953

Psychological Monographs: General and Applied

Vocabulary Knowledge and Usage among Normal and Schizophrenic Subjects[1]

Louis J. Moran

Veterans Administration Hospital, Houston, Texas

I. THE PROBLEM

The performance of schizophrenics on vocabulary tests has long been a controversial subject in psychological literature. It has been contended by some that the schizophrenic's relative lack of impairment on this test provides the clinician with a long needed "stable base line" from which to judge the degree of "mental deterioration" in other areas. By others the characteristic dispersion of other test scores around the allegedly stable vocabulary score is used as a device for differential diagnosis of psychopathology. Some have theorized that this stability of vocabulary test scores is due to the fact that word meanings represent "old learning," which is last to be impaired in psychopathology. Others have contended that this alleged stability is simply an artifact of the instruments commonly used to measure word meanings.

The present study proposes an intensive comparative analysis of the understanding and use of word meanings by matched pairs of schizophrenic patients and nonpsychiatric patients. The approach used is that suggested by Cronbach, who writes:

> . . . it appears that diagnostic vocabulary testing requires more valid instruments than are now available. It appears especially important, if diagnosis is to be accurate, that the possibility of obtaining credit without knowing the word be reduced to a minimum. Probably the most successful way to attain this is to ask several questions about each word. . . . It is hoped that a recognition of the weaknesses of present tests, and of the types of information about word knowledge they do not provide, will lead to a development of superior tests, especially to study breadth and precision of meaning (5, pp. 215-216).

The present study employs a battery of seven tests of word meaning, using a common core of 25 words in each test. The battery is designed to afford measurements, *for each word,* of the subject's (*a*) ability to define; (*b*) conceptual level of definition; (*c*) "breadth" of understanding (number of other words available to express the same meaning as that of the stimulus word); (*d*) preciseness of understanding; (*e*) ability to form concepts

---

[1] This study is a revision of a dissertation submitted in 1952 to the University of Texas in partial fulfillment of the requirements for the degree of Doctor of Philosophy. Dr. Robert R. Blake initiated, with the author, a series of studies which led directly to this study. Though Dr. Blake was in England when the present research was conducted, much of the impetus for the study should be credited to him.

To Dr. Wayne Holtzman, who directed the study, the writer is greatly indebted for his continuous constructive criticism. Drs. Seymour Fisher and Laurence McGaughran also made valuable suggestions on the design of the study. A special debt is owed to Drs. Ruth Hubbard and Holland C. Mitchell, at the Waco Veterans Administration Hospital, and to Drs. Isham Kimbell, Howard Schnur, George Granville, and Lewis Leavitt, at the Houston Veterans Administration Hospital, for their assistance in the collection of the sample of subjects used in the study.

with the words (similarities test); (*f*) conceptual level of concepts (similarities) formed with the words; (*g*) ability to communicate with the words; (*h*) ability to reason symbolically with the words; and (*i*) associations to the words. By means of such a test battery it is possible to map out, so to speak, an "area of meaning" for each word.

By administering this battery to matched pairs of schizophrenic and nonpsychiatric patients, the quantitative measures of the area of meaning around each word may be compared. Such a comparison should provide a rigorous test of the hypothesis that in schizophrenia significant distortions in word meanings take place which go undetected by present vocabulary tests.

Yacorzynski, Cameron, and Goldstein have observed distortions of word meanings by schizophrenics which relate to a lower conceptual level of definition, less precision in the understanding of word meanings, and impaired ability to communicate, to form concepts, or to reason symbolically with words. These assertions are presented more formally in section IV to derive eight specific hypotheses which are to be tested in the present study.

II. SUBJECTS

The total sample of subjects for this study consisted of an experimental group of 40 patients in a Veterans Administration hospital diagnosed as chronic paranoid schizophrenic, and a control group of 40 patients in another Veterans Administration hospital who were undergoing treatment for or recuperating from nonpsychiatric illnesses. Each schizophrenic subject was matched with a nonpsychiatric subject for age, sex, educational level, and Wechsler vocabulary raw score.

A. EXPERIMENTAL SUBJECTS

These subjects were drawn from patients who had received the diagnosis of schizophrenic reaction in a Veterans Administration hospital. This diagnosis was arrived at by means of psychiatric interviews and a clinical psychologist's report based on a psychological test battery. The final classification was made by the psychiatrist. Each patient received physical, neurological, and laboratory examinations to determine the possibility of organic complications. All patients with such complications were excluded from the sample. The diagnosis of schizophrenia covers an extremely wide variety of behavior pathology. In order to select as homogeneous a group as possible within this broad category, the entire sample consisted of those classified as schizophrenic reaction, paranoid type.

The patients in the experimental group were obtained from the "chronic" building in the VA neuropsychiatric hospital at which the sample was gathered. This building is reserved for patients who have had repeated admissions and who have remained unresponsive to treatment. No patient in the present sample had received electroshock, insulin coma, or other "drastic" treatment within a year before testing. All patients in this building who were classified as paranoid schizophrenic and who would consent to the examination were tested.

The choice of chronic paranoid schizophrenics as the experimental group was also influenced by Kasanin's work with disturbances of conceptual thinking in schizophrenia. In his classification

Kasanin found the most marked conceptual disturbances in hebephrenic schizophrenics and in "old cases of schizophrenia showing definite ideas of reference with delusions of persecution, associated with little emotional reaction to these ideas" (8, p. 47). The latter classification is broadly descriptive of the group of chronic paranoid schizophrenics in the present sample.

In all, 64 chronic paranoid schizophrenic subjects were seen. Of this number, four were too disorganized to comprehend the nature of the task, three refused to take the examination, and three refused to continue at different points of the examination. Of the remaining 54 subjects, 40 were successfully matched with control subjects.

B. Control Subjects

These subjects were drawn from the large patient population of a Veterans Administration general-medical-surgical (GM&S) hospital. Patients whose illnesses had neuropsychiatric significance or who had degenerative diseases or diseases seriously affecting the central nervous system were not included. The patients were selected from those with ailments that would be least likely to affect performance on the test battery.

There were a number of advantages in using these particular subjects as the control group whose performance on conceptual tasks was to be compared with that of schizophrenic patients who are also veterans. They had, to some degree, passed through a preselection process which made them, as a group, more comparable to the schizophrenic sample. The subjects in each group met the same minimum physical and mental requirements for service in the armed forces. Also, they passed through somewhat the same experience of exchanging civilian status for service discipline for a period of time. Variance due to institutionalization and examiner influence was minimized, since subjects in both groups were tested in similar hospital settings by the same examiner. Other variables, such as age, sex, educational level, and Wechsler vocabulary score, were directly controlled.

C. Matching

All identifying marks were removed from the Wechsler vocabulary test forms and random numbers were assigned by an assistant. The protocols were then shuffled. Each item was scored on all tests before moving to the next item. Wechsler's criteria for scoring were followed strictly. After all the Wechsler vocabulary tests had been scored, the following information about each subject was typed on separate index cards: (*a*) experimental or control, (*b*) random number, (*c*) age, (*d*) years of formal education, and (*e*) Wechsler vocabulary raw score.

Pairs were matched in terms of the information on these cards. In the matching procedure, first priority was given to the equating of Wechsler-Bellevue vocabulary raw scores for each matched pair. Educational level and age received second and third priority, respectively.

Most pairs were matched within one raw score point on the Wechsler vocabulary test. The difference in group means is 0.1 raw score point. Educational differences between pairs average one and one-half years of formal education. The mean educational level of the experimental group is 10.4 years; the control group mean is 9.5 years. The mean age of the experimental group is 41 years; the control group mean is 40 years. All subjects are males. The goodness of the matching was tested by significance of mean differences, none of which approaches statistical significance.

D. Cooperation

The work of Shakow with schizophrenics suggests that many of the "diagnostic" differences in their test performances may be due simply to their lack of cooperation with the examiner. He notes:

> In some situations, usually those of a simple type, a high level of *cooperation* is the factor which seems to be important in eliminating the differences which exist between the schizophrenic and normal levels (12, p. 387).

In order to estimate the effect of this variable, each experimental and each control subject was rated by the examiner on a six-point scale of cooperativeness. The categories of the scale are as follows:

1. Openly negative toward task. Performs very reluctantly.
2. Somewhat resistive to task. Performs with little interest.
3. Doesn't mind task. Performs with moderate interest.
4. Some enjoyment of task. Performs with more than moderate interest.
5. Task seen as a challenge. Performs with great interest.
6. Greatly motivated to make highest possible score. Very absorbed in his performance.

No attempt was made to match subjects on cooperativeness. However, little group difference in cooperativeness appears. The experimental group mean rating on cooperation is 3.6; the control group mean is 3.9. This difference of 0.3 does not approach statistical significance.

All subjects were tested and then rated for cooperativeness by the same examiner. No objective measure of the reliability of these ratings is available. Further, although the examiner made a special effort to rate the patient on attitude toward the examination rather than on the level of performance, no objective measure of the examiner's success in this effort may be provided. Another source of possible bias was the examiner's knowledge of the subject's diagnosis at the time of rating. Because of the evident subjectivity of these ratings, they are offered as only suggestive of the influence of subject cooperativeness on performance with these tasks.

III. SELECTION OF TESTS OF WORD MEANINGS

As asserted by Yacorzynski, Cronbach, and others, "knowledge of a word" implies a great deal more than the ability to define the word. The present battery follows Cronbach's suggestion that the most successful technique of measuring understanding of word meaning is to ask several questions about each word.

For the purposes of this investigation, all the test words should be familiar to all the subjects. The subjects already have been matched on their "fund of words" by means of the Wechsler vocabulary test. The specific purpose of this battery is not to tap "size" of vocabulary but to compare underlying differences in knowledge of the meanings of words that the pairs of subjects are able to define. Words with which one or both subjects are unfamiliar are lost for such a comparison and would serve no purpose in the battery. The 25 words used in the present list are among the most frequently used English words.

At the same time the words should vary in complexity, for if differences between the two groups appear, one is interested in the *level* at which they appear. One way of varying complexity is to grade the words in terms of familiarity, as is done in standard vocabulary tests. As mentioned before, however, this would tend to defeat the specific purpose of this battery. A more satisfactory way of varying complexity is to vary the "level of abstraction" of the words.

A useful distinction by Ogden and Richards on types of definitions is pertinent to our effort to select words that are very familiar but which vary in complexity. Definitions may be made either in terms of *things* or in terms of *other words* with a common referent. Thus, when

> . . . we define *words* we take another set of words which may be used with the same referent as the first, i.e., we substitute a symbol which will be better understood in a given situation. With *things*, on the other hand, no such substitution is involved. A so-called definition of a horse . . . is a statement about it enumerating properties by means of which it may be compared and distinguished from other things (10, p. 110).

Allport makes a similar distinction between types of words. He puts words into two general classes. One class

> . . . consists of words denoting objects which impinge, directly or through the surrounding media, upon our sense organs—objects which can be observed, described, manipulated, analyzed, experimented with, and generalized about (1, p. 19).

The other class of words consists of

> . . . those dealing not with objects capable of being explicitly treated, but with abstract things, such as relationships. . . . They cannot be heard, seen, handled, or pointed to with any part of our bodies (1, p. 19).

With these distinctions in mind, words may be selected which are very familiar to the subjects but which are different in "level of abstraction." The first ten words in the present list fall into Allport's first class. That is, they are words which have a "thing" referent. Each may be defined in a relatively concrete manner by a "statement about it enumerating properties. . . ." These words are, for example, *house, car, boat,* etc.

The other 15 words in the list fall into Allport's second category. They are words *without* a "thing" referent. To define these words the subject must "take another set of words which may be used with the same referent as the first, i.e., substitute a symbol which will be better understood in a given situation." These are words like *all, new, faith,* etc.

Theories concerning the "concreteness" of schizophrenic thinking suggest that the greatest difference between the schizophrenic and the control subjects might occur in the conceptual level of their use and understanding of the words in the second part of the list. White states explicitly:

> The thinking and speech of schizophrenia while of a lower order of abstraction nevertheless make use of words which we are accustomed to use to express a higher order (14, pp. 339-340).

It is entirely possible that the performance of schizophrenic subjects may not appear different from that of the control subjects on verbal tests which use simple words that are names of specific things. However, on the words that represent higher abstractions it might be that schizophrenics would continue to operate on a lower conceptual level. The words in the present list are selected so as to afford an examination of this possibility.[2]

A wide variety of techniques may be devised to test understanding of word meanings. The present selection of tests may be divided into two groups, those measuring *understanding* of the word and those measuring the word "as a symbol that can be put to some use in thinking or communication," that is, the ability to *use* the word.

The tests of understanding of word meanings are:

1. A *definition* test, which measures the ability of the subject to verbalize what he means by the word and also the conceptual level of his verbalization.
2. A *(recall) synonym* test, which measures the subject's active understanding of the breadth and preciseness of meaning of the word.
3. A *(recognition) synonym* test, which meas-

---

[2] The first ten words in the list are: house, clock, clothes, car, dirt, boat, door, food, street, garbage. The second fifteen words are: friend, big, faith, command, new, add, danger, all, strong, death, God, wise, hate, enemy, master.

TABLE 1

MEANS, STANDARD DEVIATIONS, AND SIGNIFICANCE OF MEAN DIFFERENCES BETWEEN SCHIZOPHRENIC AND CONTROL GROUPS ON TESTS OF WORD MEANINGS ($N=40$ matched pairs)

Test of Word Meanings	Mean Scores				p
	Schizophrenic		Control		
	M	*SD*	M	*SD*	
Definitions					
Adequate definitions	24.3	.3	24.5	.3	—
Abstract definitions (I+II)	14.2	5.6	14.6	5.0	—
Abstract definitions (Part I)	6.2	2.9	6.2	2.5	—
Abstract definitions (Part II)	8.3	3.2	8.1	3.0	—
Synonym (Recall)					
Percentage of correct synonyms	53.3	19.2	59.4	10.4	.05
Synonym (Recognition)					
Percentage of correct synonyms	74.8	15.6	77.2	10.9	—
Percentage of neologisms	1.0	1.4	.6	.9	—
Sentence construction					
Adequate sentences	7.9	2.6	9.7	2.6	.001
Similarities					
Adequate (Part I)	5.1	1.5	6.0	1.1	.001
Abstract (Part I)	3.3	1.8	4.2	1.5	.001
Adequate (Part II)	5.2	2.7	6.0	4.1	—
Abstract (Part II)	3.9	2.5	4.6	2.5	—
Adequate (I+II)	10.3	3.8	12.0	5.2	.001
Abstract (I+II)	7.2	4.1	8.8	3.3	.001
Percentage of adequate also scored abstract	65.7	19.7	71.9	15.8	.05
Analogies					
Number correct	6.6	2.8	7.6	2.2	.05
Word association					
Relatedness score	58.9	9.0	63.5	9.5	.001

Note: When dealing with the hypothesis that the mean of the schizophrenic group is less than the control group mean, the one-tailed test of significance is used.

creteness" of schizophrenics is reflected in their definitions of words, it is most likely to appear in the definitions they give to the last fifteen words in the list.

Description. The test itself consists of the basic 25 words to be used in all tests in the battery. The test is administered verbally, with the following instructions: "I want you to tell me what these words mean. 'House.' What does the word 'house' mean?" etc.

Subjects are urged to attempt a definition despite difficulties in verbalizing what the word means to them.

Scoring. The test is scored for the total number of abstract definitions given and for the total number of adequate definitions given. Scores on the first ten and last fifteen words in the list are tabulated separately.

Definitions are scored as abstract when they consist of a synonym of the test word (e.g., *clothes,* garments; *big,* large), or in terms of a category (e.g., *car,* vehicle; *hate,* an emotion). Definitions are scored adequate when the subject indicates clearly in his verbalization that he has some accurate understanding of the meaning of the word.

Results. It may be seen from Table 1

that the schizophrenic and control groups defined the basic 25 words equally well. The mean number of adequate definitions given by the schizophrenic and control groups is 24.3 and 24.5, respectively.

The schizophrenic group gave a mean of 14.2 abstract definitions. The control group mean of abstract definitions is 14.6. This mean difference of .4 is not significant. These data, which demonstrate no significant difference in the conceptual level of definitions given by schizophrenic subjects and by the control subjects, do not support Yacorzynski's hypothesis.

Since the last 15 words have no *thing* referent, it was felt that here the alleged inability of schizophrenics to assume the "abstract attitude" would be most clearly revealed. From Table 1 it may be seen that when results of the definitions test are divided into Part I (first ten words) and Part II (last 15 words), no significant differences between the group means appear at either level. Thus, this speculation is not supported by the data. It is notable, however, that both groups gave about 62 per cent abstract definitions to the first ten words and about 55 per cent abstract definitions to the last 15 words. It would appear that for both groups the words in the second part of the test were slightly less conducive to abstract responses than words in the first part.

B. SYNONYM (RECALL) TEST

Rationale. When a subject is asked to give as many synonyms for a word as he can, he in effect is being asked to define the word many times. His "definitions" furnish an indication of the conciseness of his concept of what the word symbolizes. He must select from numerous associations only those words that symbolize an identical concept; related but nonsynonymous words must be discarded.

According to Cameron, this is precisely what the schizophrenic is unable to do. Cameron asserts that a special characteristic of "schizophrenic language" is its extremely loose organization. ". . . the substitution of an approximate but related term or phrase for the more precise definitive term" is characteristic.

From Cameron's theoretical position it should be predicted that the schizophrenic would tend to offer as synonyms many words which, though possibly related, do not qualify as synonyms. The present test provides a measure of this alleged overinclusive and "fuzzy" understanding of word meanings in schizophrenic subjects.

HYPOTHESIS 2. *The schizophrenic's understanding of word meanings will be different from that of the control subjects in that it will be characterized to a greater extent by substitution of approximate but imprecise terms.*

Description. The test is comprised of the basic word list, in the same sequence as in the definitions test. It is administered verbally with the following instructions:

> Tell me all the words you can think of that have the same meaning as this word. For example, words having the same meaning as *little* are: small, tiny, diminutive, and so on. Now, how many words can you think of that have the same meaning as the word *house*, etc.?

Scoring.[3] The test is scored in two ways:

1. The total number of correct synonyms given.
2. The total number of words offered as synonyms that do not qualify as synonyms.

---

[3] The stimulus word *God* elicited a wide variety of esoteric words which evidently are considered synonyms for *God* by little-known sects. To avoid making arbitrary distinctions among these words, the stimulus word *God* is not scored in this test.

Results. A high percentage of correct synonyms indicates that the subject tends to be relatively exact and precise in his concept of the meaning of the test words. Conversely, a low percentage of correct synonyms indicates that the subject tends to include relatively more discrepant meanings in his concept of the meaning of the test word. Each group gave about the same mean number of total responses; schizophrenic group, 50.8, control group, 49.1. However, the schizophrenic group gave fewer correct synonyms and more incorrect synonyms than the control group.

The percentage of adequate synonyms given by the control group is 6.1 per cent greater than that of the schizophrenic group. This mean difference is significant at the .05 level. It is evident that the schizophrenics in this group are less precise in their understanding of the meanings of the test words than are the subjects in the control group. These results tend to support Cameron's hypothesis.

C. SYNONYM (RECOGNITION) TEST

Rationale. Like the recall synonym test, this test measures the breadth of word meanings in terms of the number of symbols that the subject can recognize as having a common referent. The subject must select from a given number of words only those that "belong together" and discard those that do not "fit."

According to Cameron it is just this inability to "select and eliminate" words which is found in schizophrenic language. The present test, like the recall synonym test, provides a test of Cameron's hypothesis. The two tasks differ in that the recall test requires that the patient actively select from his own associations to the word only those words with a similar referent. Full freedom of associations is permitted, and the possibility of eliciting "approximations," distorted meanings, and purely private conceptions of the word is maximized. In the recognition synonym test, the number of possibilities is limited to the number of specific alternate words given in the test, from which the subject is to select the most appropriate. The task is one of "sorting" a limited number of words, and it thus becomes a measure of the subject's ability to maintain "adequate boundaries" in his concepts under more controlled conditions. The following hypothesis is to be tested:

HYPOTHESIS 3. *Schizophrenics will be less able than the control subjects to maintain adequate boundaries in their concepts and will tend toward "overinclusion" in their understanding of word meanings.*

Description. The test is constructed around the basic word list. After each word eight choice words are listed. The number of correct synonyms varies from two to five of the eight alternates. The incorrect alternates are associated in some way with the test word but are not synonyms for it, e.g., *house,* roof; *God,* Bible. Scattered among the alternatives are a number of neologisms. The subjects are verbally instructed:

> After each capitalized word in this list are a number of words in small print. Underline each of the words in small print that has the same meaning as the word in capital letters. Do not guess.

These instructions are also printed on the test form.

Scoring.[4] Performance is scored in three ways:

1. Total number of correct synonyms underlined.

---

[4] The stimulus word *God* is not scored in this test.

2. Total number of nonsynonymous words underlined.
3. Total number of neologisms underlined.

Results. The hypothesis is tested by computing for each subject the percentage of correct synonyms among the total words underscored by that subject. The mean percentage of correct synonyms given by each group is then compared. A low percentage indicates that the subject tends to overinclude related but inexact meanings in his concept of the word meaning.

It may be seen in Table 1 that of the words underscored by the schizophrenic group, 74.8 per cent are correct synonyms. Of the words underscored by the control group, 77.2 per cent are correct synonyms. Though this mean difference of 2.4 per cent is in the direction predicted, it is not statistically significant. The hypothesis, as tested by differences in the percentage of correct synonyms given by each group, is not supported by these data.

It is notable, however, that the schizophrenic group underlined more correct synonyms, more incorrect synonyms, and more neologisms than did the control group. The mean number of all words underscored by the schizophrenic group is 77.5. The control group mean of all words underscored is 69.4. The mean difference of 8.1 is significant at the .05 level. Thus, it would appear that subjects in the schizophrenic group tended to attribute more meanings to the test words than did the control subjects.

Included among the alternate words are 13 neologisms, e.g., *demonter, euvate,* etc. The mean percentage of neologisms underscored by each group is 1.04 for the schizophrenic group and .58 for the control group. This mean difference of .45 per cent is not statistically significant.

D. SENTENCE CONSTRUCTION TEST

Rationale. The preceding three tests measure different aspects of the subject's *understanding* of certain word meanings. This test measures one aspect of his ability to use these same words to communicate. The subject is asked to "write a sentence, using these words." His ability to integrate from one to three words into a meaningful sentence provides one measure of his ability to communicate with these words. Such a measure may profitably be compared with the measures of the subject's understanding of the individual word meanings.

Current hypotheses concerning the schizophrenic's distorted understanding of word meanings have been drawn almost entirely from the schizophrenic's *use* of words in concept formation tasks, sentence completions, interpretation of proverbs, clinical situations, etc. It is interesting to speculate that the primary language disturbance in schizophrenia may be not so much in his understanding of individual word meanings as in his inability to organize individual words into a meaningful communication. A comparison of performance on the three tests of word understanding with performance on tests of ability to use the same words constitutes an examination of this possibility. Concerning the schizophrenic's ability to communicate, Cameron writes:

> The continual interchange between a given person and those around him not only develops the social character of his language and thought, but also maintains it afterward at an adequate social level. For if this organization falls below the point of intelligibility where others cannot share it, and if it cannot then be amplified by other words, gestures, signs, or demonstrations, it can no longer function in communication. As we shall see, that is just what happens in schizophrenic disorganization (4, p. 52).

The present test affords a test of the following hypothesis:

HYPOTHESIS 4. *The schizophrenic's ability to integrate words into meaningful sentences will be impaired, relative to the performance of the control subjects.*

Description. The test consists of the basic 25 words. The subject is asked to construct a sentence, using these words. Difficulty of the task is varied by increasing the number of words which the subject must integrate into a sentence, beginning with one word and building up to three words.

The words are printed on individual cards. The cards are presented to the subject, and he is asked to "write a sentence, using these words."

Scoring. Sentences are scored on a pass-or-fail basis, mainly on the criteria furnished by Wells and Ruesch (13). In order to receive credit the sentence must: (*a*) state something in accord with fact and in general "make sense," (*b*) be grammatically correct, though there may be an occasional waiver on syntax (such as singular verb and plural subject), and (*c*) contain not more than two coordinate clauses. A sentence may be declarative, interrogatory, or imperative, and the critical words may be compounded. A critical word may be changed as to number, tense, or case, but not as to part of speech, e.g., loud for loudly or loudness.

In setting up a scoring key, based on 20 subjects not included in the sample, certain additional scoring principles were made more explicit. A sentence is scored as incorrect if it (*a*) makes two statements not logically related; (*b*) fails to express a complete thought; (*c*) is so ambiguous that it defies interpretation; or (*d*) uses a stimulus word as a proper name.

Inappropriate punctuation is not scored. Certain implicit words, as *I, the,* etc., may be omitted if the meaning of the sentence is otherwise clear. Somewhat inappropriate use of nonstimulus words is permitted. A sentence which contains two statements that appear logically related is scored correct, e.g., "the door was open and the food spoiled."

Results. The schizophrenic group constructed a mean of 7.9 adequate sentences, as opposed to a control group mean of 9.7 adequate sentences. As indicated in Table 1, this mean difference of 1.8 is significant at the .001 level. The hypothesis is supported by these data.

The possibility was raised earlier that the primary language disturbance in schizophrenia might lie not so much in the schizophrenic's understanding of individual word meanings as in his ability to organize individual words into a meaningful communication. It is evident from the results of this study that the schizophrenic's language disturbance is manifested both in his understanding of individual word meanings and in his ability to communicate with these words.

E. SIMILARITIES TEST

Rationale. This is essentially a task in verbal concept formation. The subject is required to verbalize a conceptual relationship among a group of words. When the subject is asked, "In what way are these words alike?" he is asked to place them in a single category, to make a broad generalization about the words. According to Goldstein, schizophrenics have lost this "categorical attitude."

> . . . in their language there is an absence of generic words which signify categories or classes. Or, when they do use such words, analysis shows that for the patient they do not correspond to generalizations (6, p. 24).

According to Goldstein, then, the schizophrenic in the present sample should be less adequate than normal subjects in the ability to extract a general similarity from among a number of different words. This alleged inadequacy also should be reflected in a relative inability to derive a basic similarity by means of placing all words into a single, generic category. Performance on the similarities task makes possible a test of the following hypotheses:

HYPOTHESIS 5. *Schizophrenics will be less able than the control subjects to form verbal concepts with words they can define equally well.*

HYPOTHESIS 6. *The conceptual level of verbal concepts formed by schizophrenic subjects will be less "abstract" than that of the control subjects.*

Description. The test is composed of groups of two to four words which may be "categorized" under a single concept. Words in the first seven items refer to *things.* These words may be "categorized" or they may be related to each other in descriptive terms. Thus, *clock* and *ruler* may "belong together" under the generic term "measurement," or because they "both have numbers on them." Words in the second ten items have no *thing* referent and most must be categorized in abstract terms to be adequate.

Administration is verbal, with the following directions:

> Now I am going to read you some groups of words. I would like you to tell me what the words in the group have in common; that is, in what way they are alike.

Scoring. Responses in which the cluster of words is grouped into an appropriate category are scored "abstract." Responses which do not place the words in a category but describe a meaningful similarity among the words are scored "adequate." Scores on the first and second parts of the test are tabulated separately.

Results. The schizophrenic group gives a mean of 10.3 adequate similarities. The control group mean is 12.0. From Table 1 it may be seen that this mean difference of 1.7 is significant at the .001 level. It is evident that the schizophrenics, as a group, are less able than the control subjects to form verbal concepts with words that the subjects in each group define equally well. These data support Hypothesis 5.

To test Hypothesis 6, responses to the similarities test are dichotomized into abstract and not-abstract instead of in terms of adequacy. It may be seen in Table 1 that the schizophrenic group mean of abstract responses is 7.2. The control group mean is 8.8. The difference of 1.6 in group means is significant at the .001 level.

Another method of testing this hypothesis is by comparing the percentage of abstract responses to total adequate responses given. These mean percentages are 65.7 for the schizophrenic group and 71.9 for the control group. Table 1 indicates the mean difference of 6.2 per cent to be significant at the .05 level. Thus, it is evident not only that the schizophrenic group is less able than the control group to form adequate verbal concepts with these words, but also that the conceptual level of their adequate responses is lower than that of the control group. It was also speculated that the schizophrenic subjects would experience the greatest difficulty on the last ten items, which make a greater demand on the subject to assume the "categorical attitude" than do the first seven items. Actually, both groups appear to have found the last ten items more difficult than the first seven. Each group gave about 18 per cent fewer adequate responses and about 9 per cent fewer abstract responses to the last ten items. As on the definitions test, both groups evidently find words which have a *thing* referent easier to deal with than words which have no *thing* referent. However, no significant difference appears between groups in this respect.

F. Analogies Test

Rationale. This is one measure of how well the subject is able to reason with the words he has defined. The subject is asked to abstract a general principle from the relationship of two words and to find among another group of words two that have an identical relationship to each other. To perform successfully the subject must be able to generate hypotheses concerning the kinds of relationships that might obtain between the first two words and to discard all but the one relationship that "fits" among the possibilities afforded by the third word and one of the four alternate words.

Goldstein asserts that the abstract attitude is basic to this ability "to generalize, to abstract common properties, to plan ahead ideationally, to assume an attitude toward the 'merely possible' and to think or perform symbolically" (7, p. 19). Goldstein goes on to caution that analogies may sometimes be solved by roundabout "concrete" ways. However, if an impairment of abstract attitude is characteristic of schizophrenia, it seems reasonable to assume that it would appear in a comparison of the performance of 40 matched pairs of schizophrenic and control subjects on an eleven-item analogies test. The present task is considered a fair test of the following hypothesis:

Hypothesis 7. *Schizophrenic subjects will be less adequate than the control subjects in the ability to reason symbolically with words they can define.*

Description. The test consists of 11 analogies which incorporate words in the basic word list. Each analogy follows this pattern: *boat* is to *water* as *car* is to ——— float, land, wheels, ride. The subject is to underline the one appropriate word among the four alternatives.

The following instructions are given on the test form and are first read to the subject:

> In each of the lines in this test, the first two words fit each other in a definite way. You are to draw a line under the one word in capital letters that fits the third word in just the same way that the second word fits the first.

On the test form, two examples are given with the correct answer already underlined. The subject is asked to explain these two analogies. If the subject fails, the examiner explains the relationship. A third example analogy is given for the subject to solve himself. This is also explained by the subject before he begins the test proper.

Scoring. Total score is the number of correct answers.

Results. The mean score of the schizophrenic group is 6.6 as opposed to a control group mean of 7.6. This mean difference of one score point is significant at the .05 level. This difference tends to support the hypothesis that schizophrenics are less able than normal subjects to reason symbolically with words which each can define equally well.

G. Word Association Test

Rationale. This test is included in the battery to furnish evidence on the character of the schizophrenic's "associative links" to his words. The subject's understanding of the meaning of a word should be closely related to his associations to that word. Pillsbury and Meader describe this relationship in the following way:

The study of the development of words is a study of the associates a word has. Take the word *puss*. *Puss* is made to cover all small animals, but the meaning is dependent upon the associates and the associates determine the character and extent of the meaning. As the vocabulary grows and experiences accumulate, the meanings of the words change with them (11, p. 175). The meaning of a word is primarily the associates that it partially arouses and which give it color. When the associates that are excited change, the meaning of the word changes (11, p. 185).

Now, according to Cameron, the schizophrenic's understanding of word meanings is imprecise, overinclusive, and approximate. It seems reasonable to assume, then, that this "looseness" of understanding will be reflected in associations which are less closely related to the stimulus word than are the associations of the control subject. The associations test provides a test of the following hypothesis:

HYPOTHESIS 8. *The schizophrenic's associations to words that he can define will show less meaningful relationship to the stimulus word than will the control subject's associations.*

Description. The 25 words in the basic list are presented in the same order as in the definitions test. The subject is instructed as follows: "I am going to read you a list of ordinary words, one by one. When I say the word, answer with the first word that comes into your mind."

Scoring. Each response is given a score from 0 to 3 on the following basis:

0—subject does not produce an association, or subject produces a multi-word response.

1—subject produces a single word response which bears no meaningful relationship to the stimulus word, e.g., *faith*, "box."

2—subject produces a single word response which is very indirectly or "loosely" related to the stimulus word, e.g., *strong*, "physical."

3—subject produces a single word response which is closely related to the stimulus word. Some typical relations between stimulus and response words are:

a. use, e.g., *car*, "ride."

b. attribute, e.g., *car*, "wheels."

c. category (superordination), e.g., *car*, "vehicle."

d. synonym (coordination), e.g., *car*, "automobile."

e. example (subordination), e.g., *car*, "Ford."

f. opposite (contrast), e.g., *danger*, "safety."

Results. The responses of each subject were scored on a four-point scale of relatedness. The group mean scores were then compared as a test of the hypothesis. Table 1 indicates a mean score of 58.9 for the schizophrenic group and 63.5 for the control group. The mean difference of 4.6 is significant at the .001 level. Therefore, the hypothesis is supported by these data.

H. COOPERATION

It has been stated by Shakow that much of the schizophrenic's failure to perform well on vocabulary tests may be attributed to lack of cooperation. In the present study, the mean cooperation rating of the schizophrenic group, on a six-point scale, is 3.6. That of the control group is 3.9. The mean difference of 0.3 does not approach statistical significance. On the basis of these results, it would not appear that the significant differences between the performance of the schizophrenic group and the control group are due to differences in level of cooperation.

Significant positive correlations between cooperation rating and test performance are evident for both groups. These correlations range from .30 to .47 in the schizophrenic group and from .31 to .49 in the control group. Slight differences between groups in the degree to which cooperativeness is related to scores on specific tests are noted. However, these differences go in one direction as often as in the other direction.

It should be noted that these ratings

represent the subjective judgments of one examiner and should be interpreted with this reservation in mind. The writer, however, is aware of no other study in which schizophrenic subjects are rated on cooperativeness *in relation to* control subjects who are paired with the schizophrenics for age, sex, educational level, and vocabulary score, and who are examined under grossly comparable hospital conditions. The present results, though tentative, suggest the hypothesis that schizophrenic subjects are no less cooperative than other subjects *when tested under comparable conditions.*

V. DISCUSSION AND CONCLUSIONS

It is evident from the results of this study that the schizophrenic's understanding and use of word meanings differs in several important respects from that of the control subject.

Application of the scoring criteria employed by ordinary vocabulary tests does not reveal these differences. Nor do differences appear between matched pairs of schizophrenic and control subjects when the usual scoring criteria are made more elaborate. This is demonstrated by an earlier study (9), in which each definition is placed into one of five categories of conceptual level, and also by the present study.

However, when more than one response to each word is elicited, significant distortions in the schizophrenic's understanding and use of words begin to appear. It is found that the schizophrenic is significantly less precise in his understanding of word meanings. Though he may seem to define words as well as other people, he is less able to utilize these words as conceptual instruments. Further, the schizophrenic's ability to integrate words into meaningful sentences appears impaired.

Yacorzynski has offered the hypothesis that "the reason why the vocabulary of the deteriorated individuals remains relatively unchanged is because easier methods of reaching the same end results are left to the organism even if the more difficult solutions are no longer available" (15, p. 265).

An attempt to test this hypothesis by searching for differences in conceptual level of definitions given by normal and by schizophrenic subjects yielded negative results (9). It now becomes apparent that these negative findings are due more to inadequate methodology than to lack of validity of the hypothesis. The overall results of the present study clearly demonstrate that although the schizophrenic may appear to "define" words as adequately as the control subject, he is less able than the control subject to attain the "more difficult solutions" of further differentiating the word meanings, or to use the same words in reasoning or in communicating.

The fact remains that the schizophrenic's score on a standard vocabulary test does not appear to drop significantly. Nor does the conceptual level of his definitions appear different from that of the control subject. Yacorzynski's objection that this should not be taken to mean that the subject's understanding of word meanings is unimpaired seems valid. However, the reason why this response shows no apparent impairment while other kinds of responses to the same words do show impairment, remains to be explained.

It does not appear necessary to evoke a concept of differences in "age" of each learned response to account for this phenomenon. Such a concept would be difficult to defend. Why, for example, does the schizophrenic show impairment in ability to communicate with words that he can define? It does not seem a reasonable assumption that the schizophrenic has had more practice over a longer period in defining words than he has had in communicating with these words.

A more parsimonious rationale may be offered to account for the findings of the present study. Most definitions tests require simply that the subject indicate that he knows something of the meaning of the test words. In the synonym tests, similarities, analogies, and sentence construction tests, the subject is required to relate one word with other words in combinations of varying complexity. These tests, then, make a greater demand on differentiation *between* words than does the definitions test. Thus, in his understanding of word meanings the schizophrenic may be said to fail in finer differentiations of word meanings, although he may still be able to identify individual words.

This rationale is in close agreement with Goldstein's (7) concept of "dedifferentiation of function." Goldstein asserts that in behavioral pathology those performances which are most differentiated are lost first, while the less differentiated ones are preserved longer. Further, "dedifferentiation of language will follow the same rules which determine dedifferentiation of performance in general" (7, p. 29). Such subjects, according to Goldstein, will fail equally in familiar situations and in new ones. It is not the varying familiarity with tasks which accounts for differences in performance levels, but the varying demands of each task in terms of the differentiations required.

The discussion has stressed primarily the differences between the schizophrenic and the control subjects. Something should be said of the similarities. The group differences, though statistically significant, do not serve to separate cleanly all members of each group.

The degree of "overlap" in the performance of subjects in these two groups may be illustrated in the following way. The analogies test mean score of the schizophrenic group is 6.6. This figure corresponds to a z (or sigma) score of .36 in the control group distribution. In terms of the normal curve, then, 36 per cent of the scores of the control subjects fall *below* the mean of the schizophrenic group. In a similar calculation it may be found that 33 per cent of the schizophrenics' scores on this test fall *above* the mean of the control group. On all tests, from 25 to 38 per cent of the schizophrenics have scores higher than the control group mean. And from 25 to 36 per cent of the control group subjects scored below the schizophrenic group mean on these tests.

It is obvious, then, that the group differences reported cannot be taken as uniquely characteristic of schizophrenic individuals. These behaviors are found in all subjects. The differences are a matter of degree.

It does seem reasonable to assume, however, in the light of the group differences found, that the impact of schizophrenia results in an impairment of the individual's understanding of word meanings. This impairment, however, is so subtle and apparently so relative to the subject's premorbid ability, that a test battery such as the one used in this study certainly does not suffice to "diagnose" schizophrenia.

VI. SUMMARY

This study provides an intensive comparative analysis of the understanding and use of word meanings by schizophrenic patients and by nonpsychiatric patients. Eight hypotheses concerning alleged specific distortions in the schizophrenic's understanding of word meanings are tested. These hypotheses are derived primarily from the theorizing of Yacorzynski, Cameron, and Goldstein.

The performance of 40 matched pairs of schizophrenic and control subjects on a special battery of verbal tests is compared. Each pair is matched for Wechsler-Bellevue vocabulary raw score, age, sex, and educational level. Subjects in both groups are rated on cooperation.

All tests in the special battery are built around a common core of 25 familiar words. The battery includes tests of definitions, synonyms, sentence construction, similarities, analogies, and word association.

When the schizophrenic group's performances on this battery are compared with the control group's, the following differences become evident: (*a*) The schizophrenic is significantly less precise in his understanding of word meanings. This is most evident in his relative failure to differentiate between word meanings. (*b*) The schizophrenic is less able to use his words as conceptual instruments, e.g., his ability to form verbal concepts and to reason in analogies tests appears impaired. (*c*) The schizophrenic's ability to integrate words into meaningful communications appears impaired, though he may "define" the same words in the same way that the control subject does.

These results support the theories of Yacorzynski, Cameron, and Goldstein concerning the nature of the schizophrenic's understanding and use of word meanings. It is noted, however, that the differences in the schizophrenic's understanding of word meanings are not large enough to be diagnostic. Despite group differences, there remains an extensive "overlap" in group performances of the schizophrenic and the nonpsychiatric patients.

Cooperation ratings correlate about .40 with performance level. No significant differences between the groups in level of cooperation are evident.

REFERENCES

1. ALLPORT, F. H. Institutional behavior. In L. J. Lee (Ed.), *The language of wisdom and folly*. New York: Harper, 1949. Pp. 15-20.
2. CAMERON, N. Reasoning, regression and communication in schizophrenics. *Psychol. Monogr.*, 1938, 50, No. 1 (Whole No. 221).
3. CAMERON, N. Experimental analysis of schizophrenic thinking. In J. S. Kasanin (Ed.), *Language and thought in schizophrenia*. Berkeley: Univer. of California Press, 1944. Pp. 50-64.
4. CAMERON, N. *The psychology of behavior disorders: a biosocial interpretation*. Boston: Houghton Mifflin, 1947.
5. CRONBACH, L. J. An analysis of techniques for diagnostic vocabulary testing. *J. educ. Res.*, 1942, 36, 206-217.
6. GOLDSTEIN, K. Methodological approach to the study of schizophrenic thought disorder. In J. S. Kasanin (Ed.), *Language and thought in schizophrenia*. Berkeley: Univer. of California Press, 1944. Pp. 17-40.
7. GOLDSTEIN, K. *Language and language disturbances*. New York: Grune & Stratton, 1948.
8. KASANIN, J. S. The disturbance of conceptual thinking in schizophrenia. In J. S. Kasanin (Ed.), *Language and thought in schizophrenia*. Berkeley: Univer. of California Press, 1944. Pp. 41-49.
9. MORAN, L. J., MORAN, FRANCES A., & BLAKE, R. R. An investigation of the vocabulary performance of schizophrenics: II. Conceptual level of definitions. *J. genet. Psychol.*, 1952, 80, 107-132.

10. OGDEN, C. K., & RICHARDS, I. A. *The meaning of meaning: a study of the influence of language upon thought and of the science of symbolism.* New York: Harcourt, Brace, 1918.

11. PILLSBURY, W. B., & MEADER, C. L. *The psychology of language.* New York: Appleton, 1928.

12. SHAKOW, D. Some psychological features of schizophrenia. In M. L. Reymert (Ed.), *Feelings and emotions: the Mooseheart symposium.* New York: McGraw-Hill, 1950. Pp. 383-390.

handbook. New York: Psychol. Corp., 1945.

13. WELLS, F. L., & RUESCH, J. *Mental examiner's*

14. WHITE, W. A. The language of schizophrenia. *Proc. Ass. Res. nerv. ment. Dis.,* 1928, **5**, 323-343.

15. YACORZYNSKI, G. K. An evaluation of the postulates underlying the Babcock Deterioration Test. *Psychol. Rev.,* 1941, **48**, 261-267.

(Accepted for publication August 3, 1953)

Vol. 67, No. 21 Whole No. 371, 1953

Psychological Monographs: General and Applied

The Use of Modified Block Designs in the Evaluation and Training of the Brain-Injured

Marion Reissenweber[1]

Institute for the Crippled and Disabled, New York, N.Y.

This report describes the development and present use at the Institute for the Crippled and Disabled, New York City, of modified Block Design materials and methods as an aid in evaluating and in training some disturbances in the visual perception of form, size, space, and organization incident to brain injury but not clearly outlined as clinically describable visual difficulties. The modifications in materials and methods are the outgrowth of the writer's personal need for such changes and her subsequent observation of similar needs in brain-injured patients receiving psychological retraining in visual perception and organization. The account is presented for whatever practical or suggestive help it may offer.

No attempt will be made in this report to discuss the theories of perception or the relative importance of the factors contributing to disturbances in visual perception and organization. With reference to such considerations, it may be empirically stated, however, that disturbances in visual perception and organization vary greatly in terms of the extent and nature of brain injury, as well as in terms of all the inherent and environmental factors which may influence an individual's response; that severe disturbances in the perception of form and space and in visual organization can be present in an individual who shows good ability in other aspects of thinking; and that when suitable methods of training are employed, some improvement in practical function may be expected in more instances than would seem possible on the basis of the Wechsler-Bellevue Performance Scale results in this series of patients.

Statement of the Problem

The personal need for increased simplification and concreteness in the Block Design Cards of the Kohs (11) and Wechsler (16) series first became apparent when attempting to use these Design Cards as an aid in self-retraining after brain injury resulting in severe disturbances in visual perception and organization. Since at that time the writer was not familiar with the reports of Goldstein and Scheerer (6), it was of interest to discover later that the steps suggested by them for progressively simplifying the Design Cards in their own series (7) were among the writer's first felt needs. Very gradually it became possible to recognize some of the specific difficulties presented by the Kohs and Wechsler materials. Outstanding among these were the absence of block divisions on the Design Cards; the actual size difference between 1-inch blocks and Design Cards drawn to the scale of ½ inch to one block; the completely different appearance of a two-dimensional Design Card and its three-dimensional block representation; the flat appearance of both Design Cards and block representations; and the presence of margins on the Design Cards.

---

[1] Grateful acknowledgment is made to Dr. Arthur L. Drew, Jr., Department of Neurology, University of Michigan Medical School, Ann Arbor, Michigan, for his counsel and encouragement in the preparation of this manuscript.

From an evaluation of these observations there seemed reason to believe that, with modifications, block design construction could be more useful as a diagnostic aid in recognizing some disturbances in visual perception and organization. Furthermore, as a method for training or retraining in some cases of disturbances of form-space perception and visual organization, particularly within the field of central vision, there seemed to be indications of its value. To these ends modifications were made in the blocks, the design cards, the working environment, and the use of materials.

Clinical Material

During the last three years these materials have been used at the Institute for the Crippled and Disabled with brain-injured patients who have been referred either for qualitative psychological evaluation or for psychological retraining in visual perception and organization. Within the group of 50 patients with whom these materials have been used, there have been no selective criteria except a medical record of brain injury and evidence of resulting disturbances in visual perception and organization. The etiology of brain damage included birth injuries, war injuries, other accidental injuries to the head, encephalitis, brain surgery, and cerebral vascular accidents. The disturbances in visual perception and organization were primarily variously combined distortions and displacements of form, space, size, and position rather than clearly recognized visual agnosias, alexias, or aphasias. The patients ranged in age from 8 to 46 years. Verbal intelligence as estimated by the Wechsler-Bellevue Intelligence Scale ranged from 62 to 126. On the Wechsler-Bellevue Performance Scale, achievement ranged from 38 to 103, with all but eight patients at levels below 90. Discrepancies between verbal and performance scale achievements ranged from 6 to 73 points, with 50 per cent of the group averaging a difference of 30 points. A further publication will contain a discussion of the correlation of these findings with the results on our modified block design evaluation.

Recognition of Limitations in Standard Materials

In groping efforts to meet the personal feeling that the Block Design Cards of the Kohs and Wechsler series needed to be simpler and more concrete, the standard Design Cards of the Wechsler series were first redrawn from block models known to be correctly constructed. The actual block divisions that were thus exposed provided one tangible clue to visual orientation and organization. For example, in constructing Design Card 3 of this series (16), placing the two solid white blocks so that the lower right-hand corner of one block touched the upper left-hand corner of the other provided a spatial orientation for the remaining two blocks (Fig. 1a). The positions of these remaining blocks were then determined by noting that when correctly placed (as verified by a friend) the diagonal division on each of these blocks extended from the upper left-hand corner to the lower right-hand corner and that the red part of each block was turned to the inside. When so placed it was noted that the white tip of each block was to the outside. Since the writer found that this diagonal dividing line could be established by verbalizing its terminal

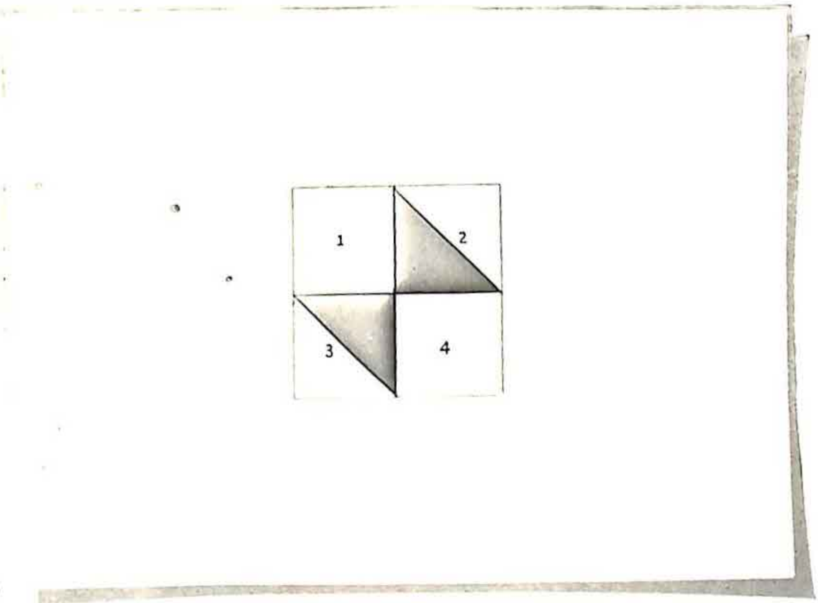


FIG. 1a. Use of Numbering and Indented Outline to Clarify Block Construction of Design Card 3 of Wechsler-Bellevue Performance Scale, Form 1 (Courtesy of Psychological Corporation, New York).

points and could be checked by kinesthetically tracing that direction, she considered this dividing line to be a much simpler and more reliable clue to block placement than were the white corners or the "bowknot" configuration of the design, which at that time could not be fused and had no wholeness. Since the block divisions and the diagonal lines seemed to provide the most personally useful clues to visual orientation and organization, the first drawings which the writer made of these standard designs were made with lines that indicated both the divisions between blocks and the lines that divided blocks diagonally. It should be noted here that this use of lines was a simple attempt to establish tangible orientation clues by which a Design Card could be broken down into manageable parts. This use of lines should not be confused with, nor considered comparable to, the use of lines as described by Goldstein and Scheerer (6). On these first drawings the block representations were numbered as another help in visualizing the breakdown of the whole design into parts (Fig. 1a). The outer lines of the design were then deepened by tracing them with the pointed end of a leather tooling instrument—a device which for the writer served first to establish a kinesthetic appreciation of the continuous outline of the figure against a background. Later this slightly indented outline was sufficient to raise the figure enough to provide a visual clue which increased the ease of both kinesthetic and visual tracing of the design, and which from some positions made it fleetingly possible to experience something akin to a foreground-background relationship.

From these drawings it was possible to draw the designs on an expanded scale that corresponded exactly to the size of the blocks. On these larger drawings both the whole and half-block divisions were indicated by lines. The outlines of the designs were then deepened with the leather tooler and, in addition, the figure itself was painted with a coat of colorless nail polish, which had little gloss, but which helped further to bring the figure up from its background (Fig. 1b). With these large cards it was possible to use the direct hand-kinesthetic method to trace some of the outlines and to help in determining the directional trend of stripes.

Even when the Design Cards were drawn to an expanded scale and tooled with line divisions to indicate their organization, construction continued to be laborious and confusing because the completed three-dimensional block design bore little, if any, resemblance to the Design Card (Fig. 2). Correctness of construction could be established only by a slow and systematic tracing of each part on both the block design and the Design Card. There seemed to be no visual wholeness of configuration on either one. Furthermore, the appearance of extreme flatness of the Design Card was complicated by directional confusions that seemed to alter the length and proximity of lines as well as the size and position of solid parts (Fig. 2b). The same disconcerting flatness distorted clues to puzzles of the jigsaw type and destroyed completely the ability to appreciate cartoons. For example, the pieces of a jigsaw puzzle when seen from a sitting position lacked individual character and conveyed only a general color impression. From neither a sitting nor a standing position was it possible to appreciate the picture as a whole. It was noted consistently that the outline of individual pieces provided no useful clues for construction. Such puzzles could be successfully completed, however, when the individual pieces could be viewed from positions that made it possible to see and feel their depth and contour.

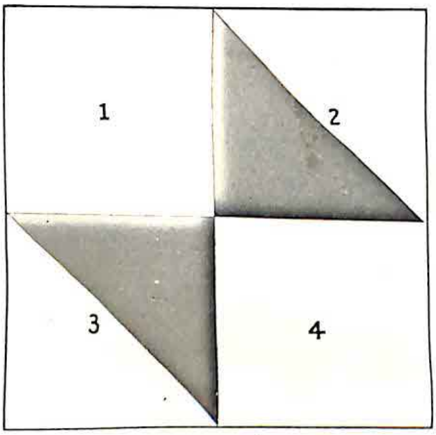


FIG. 1b. Further Simplification of Card 3 by Enlarging Drawing to Correspond Exactly with Size of Block Construction.

After some experimentation with various positions and lightings it was found that when the Design Card was placed on an inclined surface, it was possible to get some relief from the flatness that prevented any figure-background relationship. With further experimentation it was possible to devise an incline on which both the Design Card and the block representation of the design seemed to approach a three-dimensional experience (Fig. 3).

The improvement thus effected by the use of an inclined surface made it possible to localize another source of difficulty—namely, the inescapable influence of contours and surfaces of objects within the visual range. The edges of the table, the space between table and wall, the outline of window and door casings, the angulations of a book or box within the visual field all influenced judgments of shape and size unpredictably. Relative to this it should be noted that the contours of the visual world beyond the boundaries of immediate visual fixation can be monumental obstacles which are capable of increasing spasticity and disrupting visual organization to a degree which is difficult for a person without impairment in visual perception to understand.

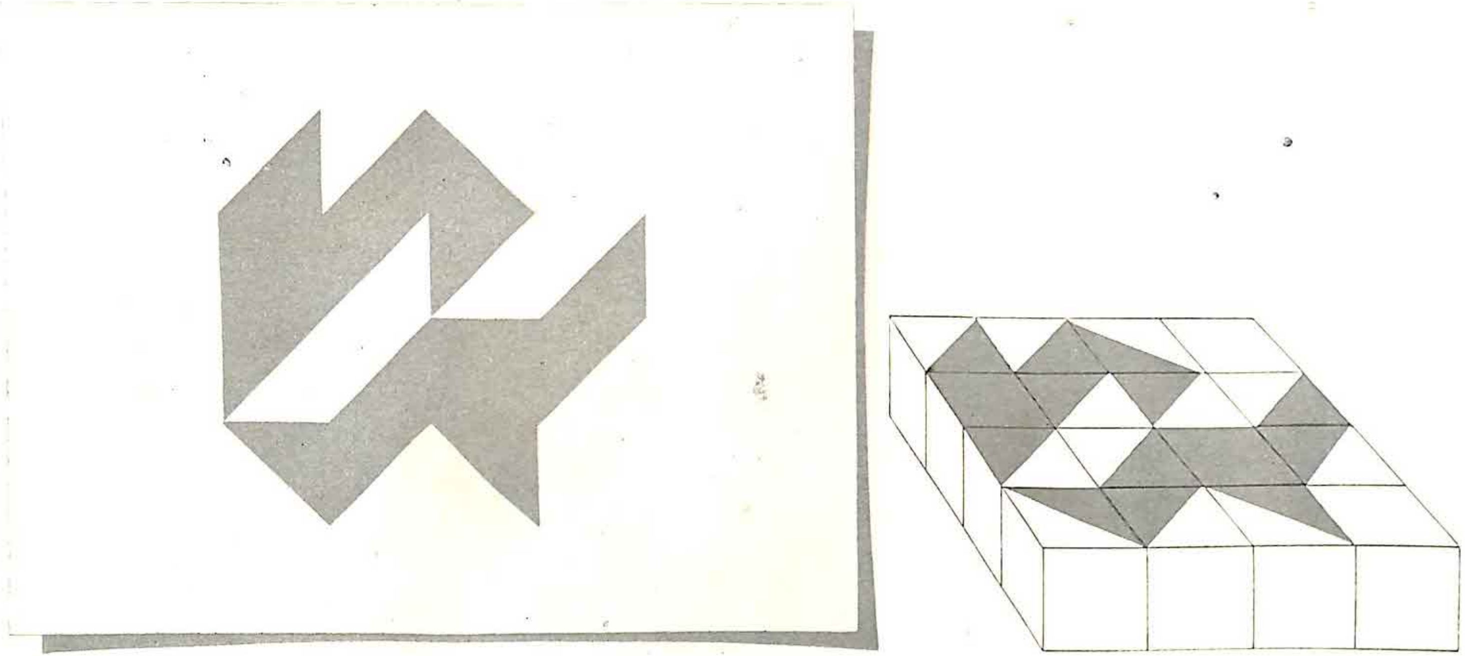


Fig. 2. Comparison of Standard Two-Dimensional Drawing of Design 17 of Kohs Block Design Series (Courtesy of C. H. Stoelting Co., Chicago) (Fig. 2a) and Its Three-Dimensional Counterpart Drawn to Same Scale (Fig. 2b). In actual use, individually perceived differences are frequently greater because standard blocks are twice the size of this standardly scaled drawing.

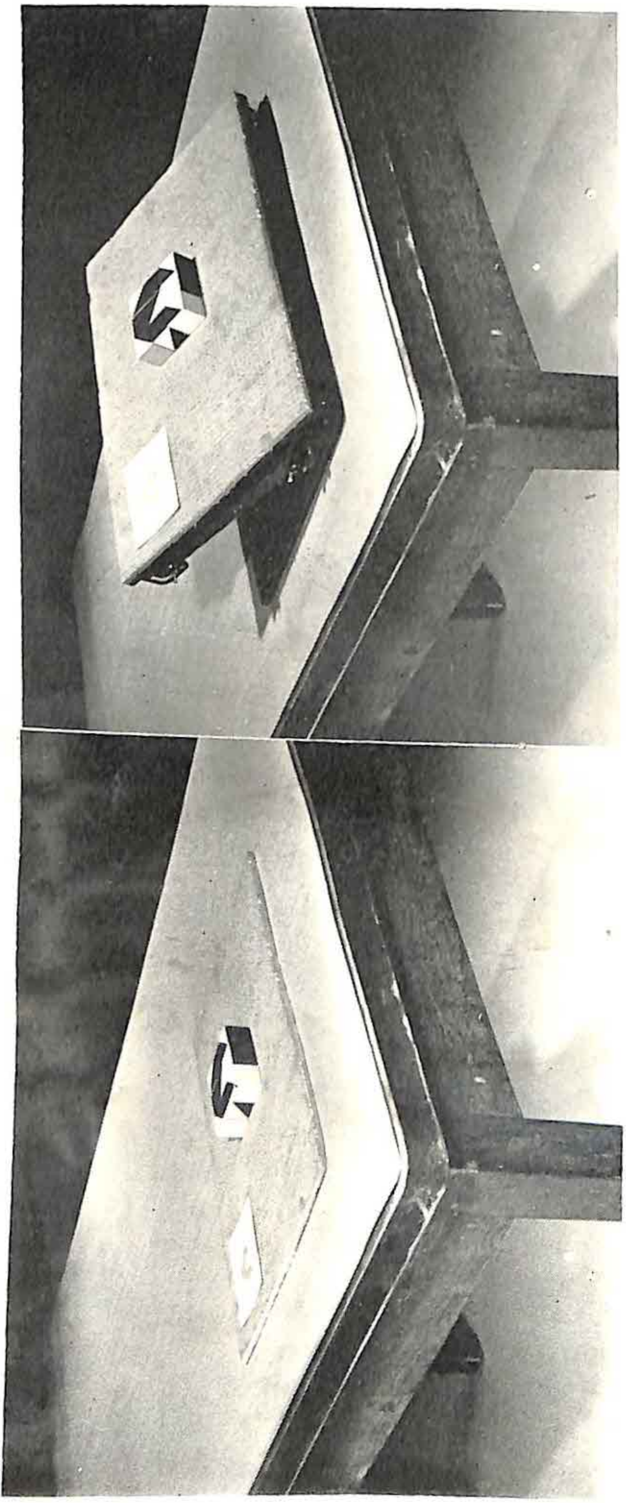


Fig. 3. Comparison of Card 11 of Kohs Block Design Series (Courtesy of C. H. Stoelting Co., Chicago) and Its Block Representation as They Appear on a Flat and an Inclined Surface.

When the physical environment was arranged in a way that ruled out gross disturbances of this sort, it was possible to see that at least one other obstacle remained. This was the white margins on the Design Cards.

With the help of these simplifications it became possible, through systematic practice and verbal analysis of the organization of each correctly completed design, to construct from memory each of the Kohs and Wechsler Design Cards from any one of four positions. Had the writer been familiar with the Goldstein-Scheerer Design Cards with lines (7), early attempts at retraining would undoubtedly have been easier. In the absence of this help, adaptations as previously described were made for each of the Kohs and Wechsler cards, and self-practice was directed toward the ability to construct the designs from memory. For example, in Card 7 of the Wechsler-Bellevue Scale, Form 1 (Fig. 4), the visual organization of this design would be fixed in memory by verbalizing as follows: The central unit is a white diamond made up of two two-block white triangles. The center part of each side is made up of a two-block red triangle placed with its point to the inside and touching the adjacent white point of the center diamond. Each corner is made with one block placed so that the red tip is to the outside. It was noted with concern, however, that although this device was personally a real help in reproducing the designs in the practice series, the ability to construct these designs did not increase skill in organizing and constructing a design not previously studied.[2]

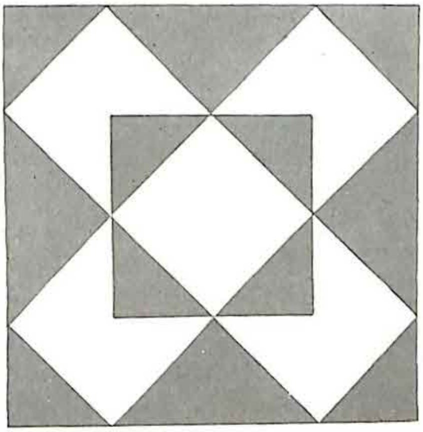


FIG. 4. Design Card 7 of Wechsler-Bellevue Performance Scale, Form 1 (Courtesy of Psychological Corporation, New York).

Subsequently, in observing the use of Kohs Design Cards by patients who had been receiving retraining in perceptual organization in this situation, there was further evidence to confirm the observation that persons with good intellectual potential and severe perceptual disturbances could master the accurate construction of the Design Cards as a series of concrete tasks without any measureable improvement in their ability to analyze or construct unfamiliar designs.

Modifications in Blocks and Design Cards

Modifications in the blocks consisted in substituting 16 one-inch wooden blocks finished in black and white enamel for the standard cubes of red, white, blue, and yellow. Two sides of each block are black, two are white, and two are diagonally divided and finished half in black and half in white. For convenience these are boxed in sets of 16. In actual use this number is frequently supplemented to make a working set of 25. Another set of blocks consists of 16 one-inch wooden cubes that have been sawed diagonally to make 32 half-blocks.

---

[2] It should be noted that verbalization of this sort is not suggested as a uniformly helpful device. Subsequent experience has shown that with some patients it is a hindrance; with others it appears to be a help.

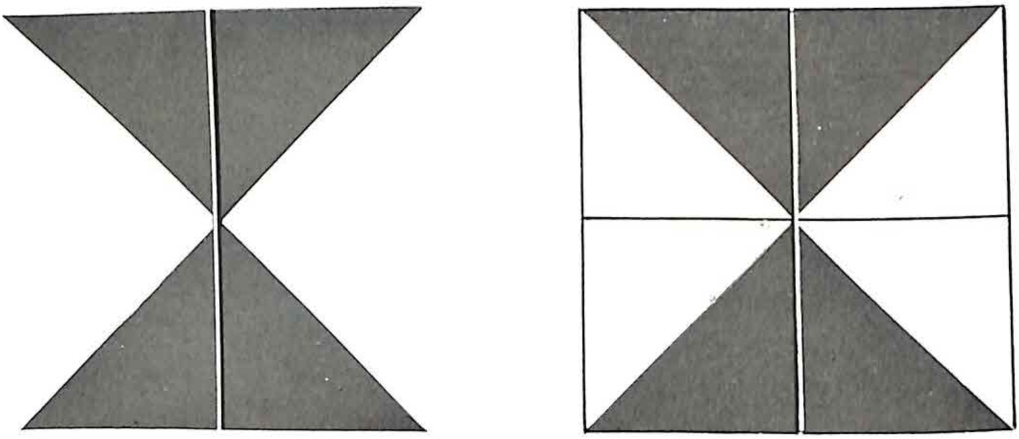


FIG. 5. Use of Black Half-Blocks to Build a Figure (Fig. 5a) Apart from Its Background (Fig. 5b).

Half of these are white, half are black. With these it is possible to build a figure apart from its block background (Fig. 5).

This change from the standard red-white, blue-yellow blocks used by Kohs (11), Wechsler (16), and Goldstein and Scheerer (7) to black and white was made in order to provide the sharpest and strongest contrast between figure and background, to eliminate the possibility of failure or confusion because of individual physical or emotional response to color (13), and to reduce the possibility of failure because of the presence of certain brain lesions that may make it difficult, if not impossible, for the eye to pick up small color differences (9, pp. 55-56; 10, pp. 25 and 125; 14, pp. 38-39). These considerations are important in this situation since routine medical procedures do not include a detailed differential diagnosis of brain-injured persons, nor do the routine psychological examinations rule out the possibility of color blindness or of color deviations related to other pathological conditions. Since such deviations alone do not bar a patient from acceptance for psychological retraining, efforts have been made to provide materials which present the strongest black-white contrast between figure and background and are, therefore, apt to be useful to the largest number of patients.

Since it was felt that the standard design cards were not entirely applicable to the problem at hand, original design cards in black and white were developed for use with these blocks. It should be emphasized here that these materials are not to be considered as a test. It is believed that they do serve a useful purpose as part of a qualitative evaluation of brain-injured persons, and that their evaluative usefulness continues when they are used in retraining. They are not, however, to be confused with, nor considered a substitute for, the Goldstein-Scheerer Cube Test (7). At present there are more than 50 different designs of varied size and difficulty planned to aid in recognizing and differentiating the nature of some of the disturbances in visual perception. These same designs have been applied to training and retraining in visual perception and organization. The sizes of the actual design cards vary from 1 × 1 inches to 5 × 5 inches. These variations depend upon the number of blocks required for the construction of a design, the scale used in drawing the design, and whether or

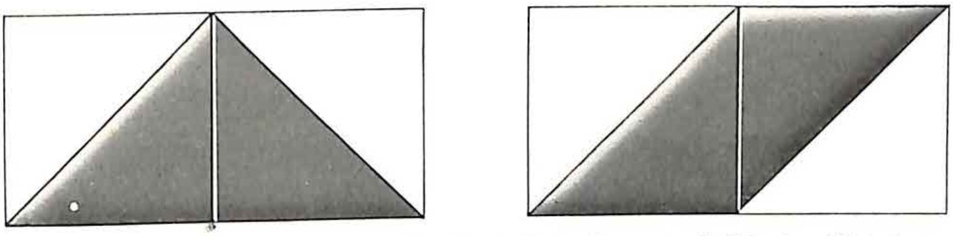


FIG. 6. Sample Design Cards for Basic 2-Block Units, Drawn to Full Scale, with Whole- and Half-Block Divisions Shown.

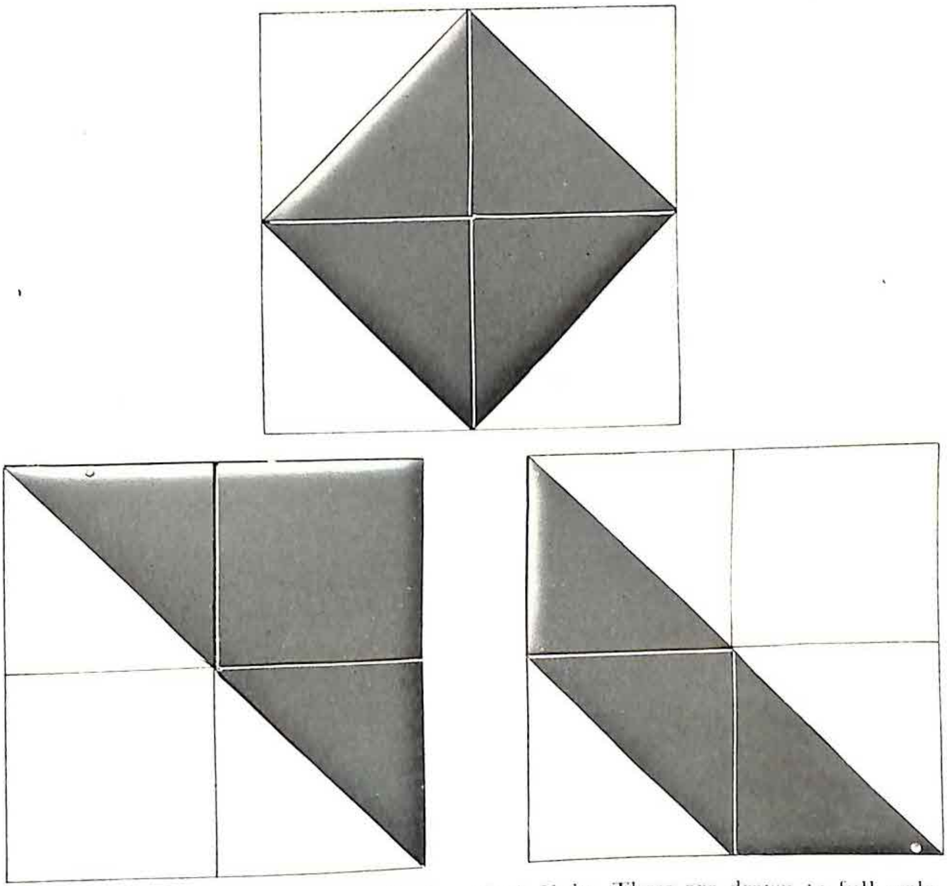


FIG. 7. Sample Design Cards for Basic 4-Block Units. These are drawn to full scale, with whole- and half-block divisions indicated.

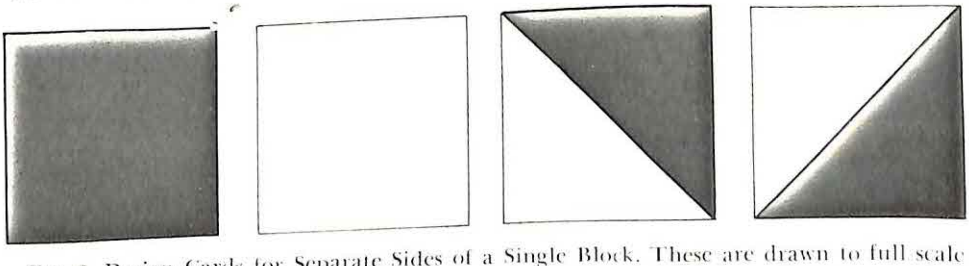


FIG. 8. Design Cards for Separate Sides of a Single Block. These are drawn to full scale so that construction can be clarified by placing half-blocks directly on top of corresponding parts of the card.

not extraneous margins are used. For example, a construction may require 25, 16, 9, 15, or 12 blocks. Several cards have been made up for 2 and 4 blocks (Fig. 6 and 7). There are also card replicas of single blocks so that when necessary it is possible to simplify a problem to a point where a single block and a single card are used (Fig. 8). In very severe cases, where it seems almost impossible for a patient to see any relationship between a three-dimensional block and its two-dimensional representation, the use of a three-dimensional cardboard model of a single block has often been found helpful. This is folded and cut in such a way

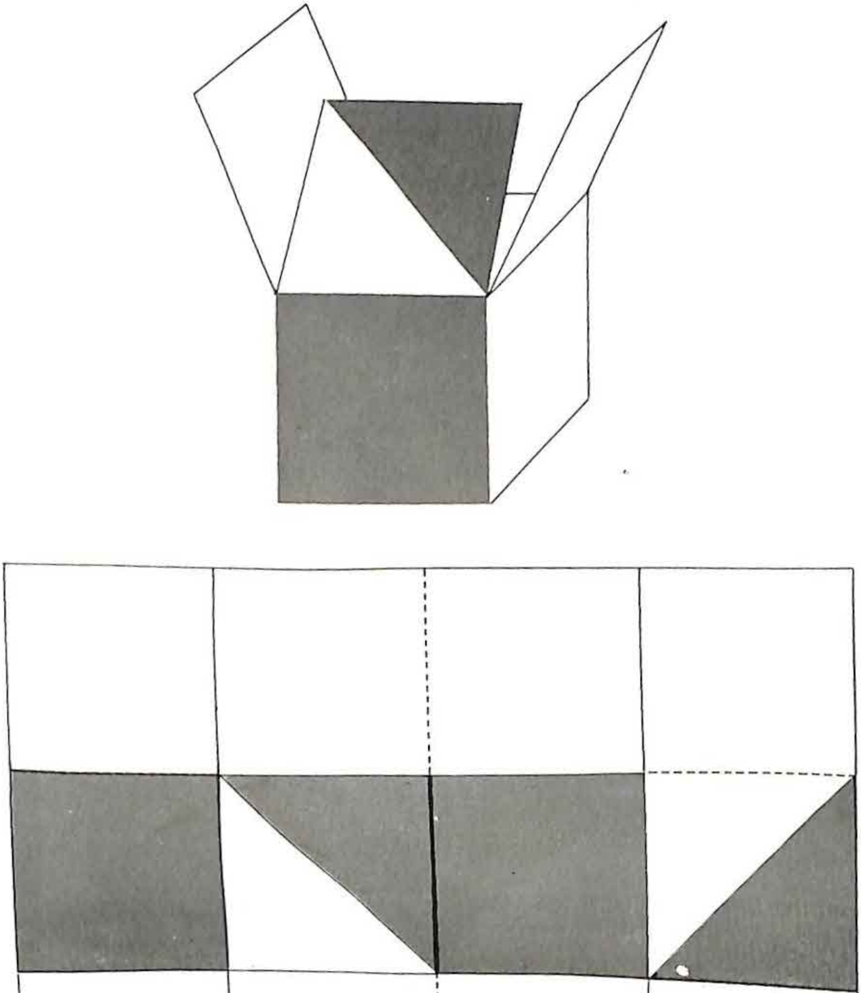


FIG. 9. Cardboard Model of Single Block which Can Be Used in Upright Position (Fig. 9a) and Can Also Be Flattened (Fig. 9b) and Masked to Show a Flat Representation of Each Side of the Block.

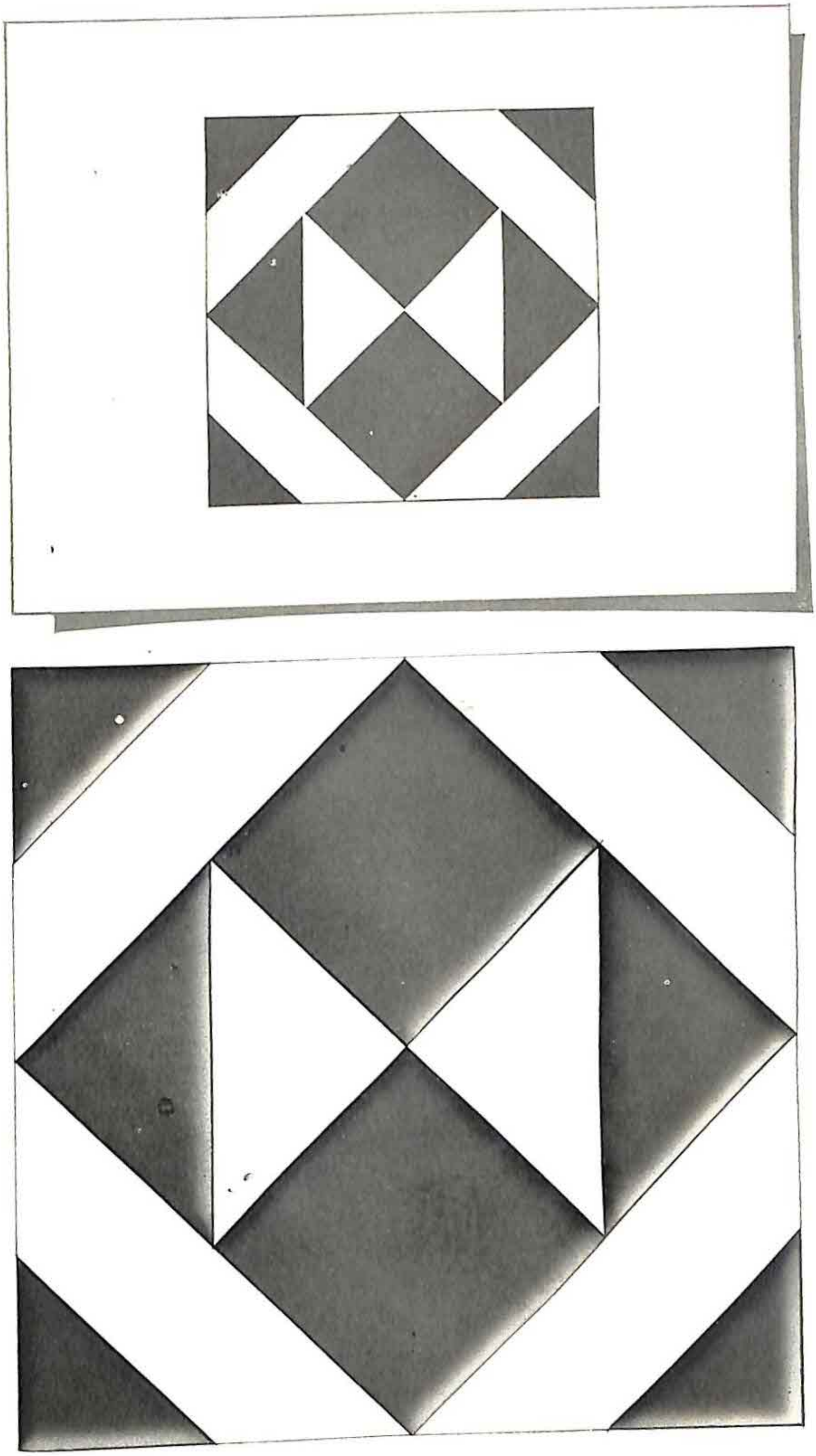


Fig. 10. Sample 16-Block Design Card, without Margins, Drawn to Actual Size of Total Configuration (Fig. 10a) as Compared with Same Design Drawn to Smaller Scale and Placed Within a Frame (Fig. 10b).

that it can stand as a one-inch cube and can also be spread out on a table. When using this in conjunction with a wooden block, each side of which can be shown comparably on the upright cardboard three-dimensional model and can also be flattened out and isolated by means of a cardboard mask which covers all but the side of the block under consideration, patients have frequently been able for the first time to visualize a relationship between a flat two-dimensional representation of a design and its three-dimensional counterpart (Fig. 9). The configuration of the total figure-backgrounds to be constructed is squares, diamonds, and rectangles. The design cards used during early stages of training are exactly the size and shape of the total figure-background to be constructed. That is, the design extends to the limits of the card without extraneous frame (Fig. 10a), thus eliminating the patient's need to find some way to cope with the very real visual obstacle that is frequently presented by the space between the frame of the design and the edge of the card (Fig. 10b). Furthermore, by making design cards in such a way that the figure and background to be constructed fill the entire card, it has been found possible to use orientation clues in terms of kinesthetic understanding—that is, top, bottom, lower right corner, etc. have concrete meaning, and block placement can actually be checked by feeling. In

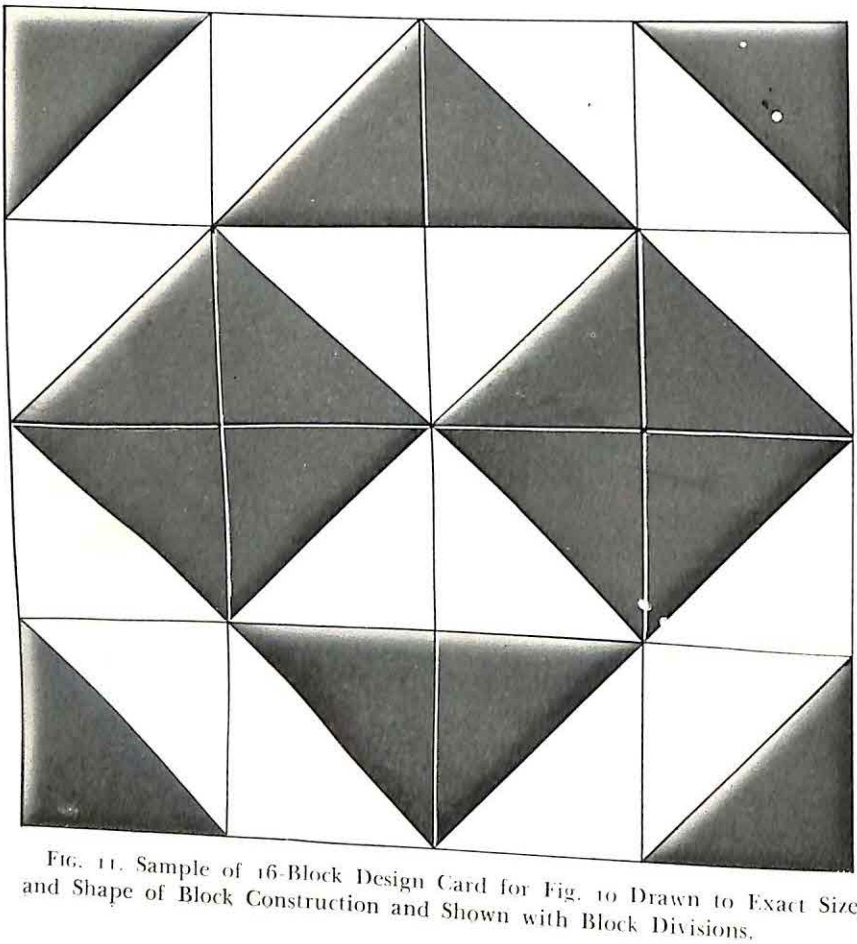


FIG. 11. Sample of 16-Block Design Card for Fig. 10 Drawn to Exact Size and Shape of Block Construction and Shown with Block Divisions.

patients who show directional confusions, the elimination of the margins which appear on the standard design cards seems to greatly facilitate training.

Some of the designs which have been found most useful at very basic levels are made up in several ways. The simplest of these are drawings similar to one of the modifications described by Goldstein and Scheerer (7, p. 54). They are drawn to the scale of one inch to a block. They are made up both with and without line divisions between the blocks and half-blocks (Fig. 11 and 10a). These same designs are also made up on a smaller scale in two forms, with and without line divisions. Neither form provides for any margin beyond the design itself. As previously noted, it has sometimes been found useful during early stages of training to reinforce the outline of the total configuration (Fig. 10a). Because of this the modified materials include some design cards that show an indented outline of the figure which is sufficiently clean-cut to permit patients whose hand coordination is adequate to trace the figure.

For later stages in training, designs consisting of a figure and background within a frame are placed on cards of various sizes and shapes. These do provide a margin between the frame of the design and the edge of the card. These are sometimes a square (Fig. 10b), a diamond, or a rectangle on a square or rectangular card, or they may be a square frame on a diamond-shaped card (Fig. 12). Further variations may be obtained by presenting the cards in rotated positions. The levels of difficulty of these designs are comparable in terms of physical form to the Kohs series, Cards 1-16, and all of the Wechsler cards, except for such simplifications as may result from the use of black and white instead of red, white, blue, and yellow.[3]

Some of the design cards that have appeared to be most useful have been drawn to various scales and placed on cards of various sizes. On these there are no frame lines to indicate where the unit to be constructed ends (Fig. 13). This level of difficulty is believed to be comparable to Card 17 in the Kohs series. The increased demands which this type of construction places upon the form and space orientation of a patient, as well as upon his visual organization and his ability to handle visual abstractions, would be apparent to the most casual observer. These added difficulties range from a gross loss of orientation and organization resulting from the fluid merging of design-background and extraneous margin to a recurrence of confusions in position and direction encountered by the patient at earlier stages in training. "You really did things to me with this one" is not an uncommon comment.

The design cards that have been developed for use with these modified blocks also include a series in which the design, without frame boundaries, is drawn upon a structured background. Some of these backgrounds are fairly regular and dependable in form and color intensity (Fig. 14a and b). Others are irregular but fairly even in intensity. Still others have scattered points of distraction (Fig. 14c). By rotating the positions in which these design cards are presented, it is often possible to gain additional clues to field abnormalities as well as to the nature of other existing disturb-

---

[3] "Simplifications" is used in the plural advisedly because I do not think that one can say that only a single simplification is effected for any individual patient.

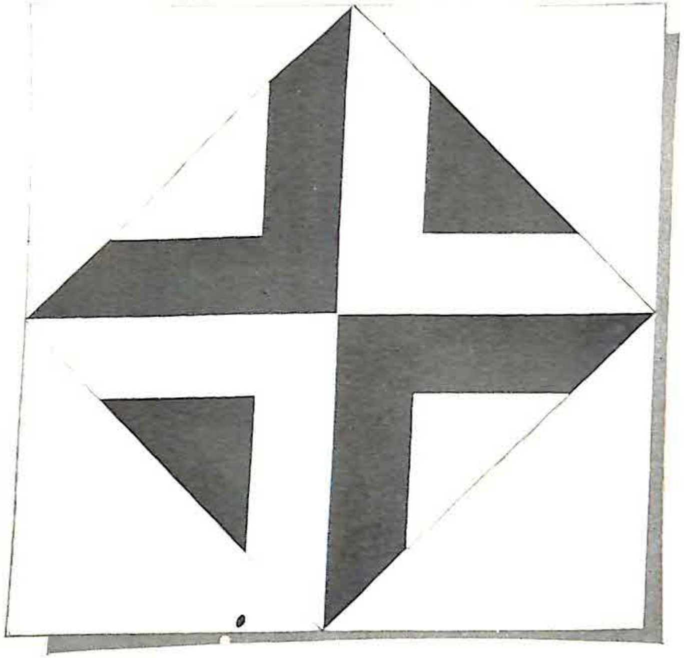


FIG. 12a. Sample Design Card for 16-Block Diamond Drawn to Small Scale and Placed within a Frame on a Square Card. The absence of block divisions and the space between the frame and the edges of the card make further demands upon visual perception and organization.

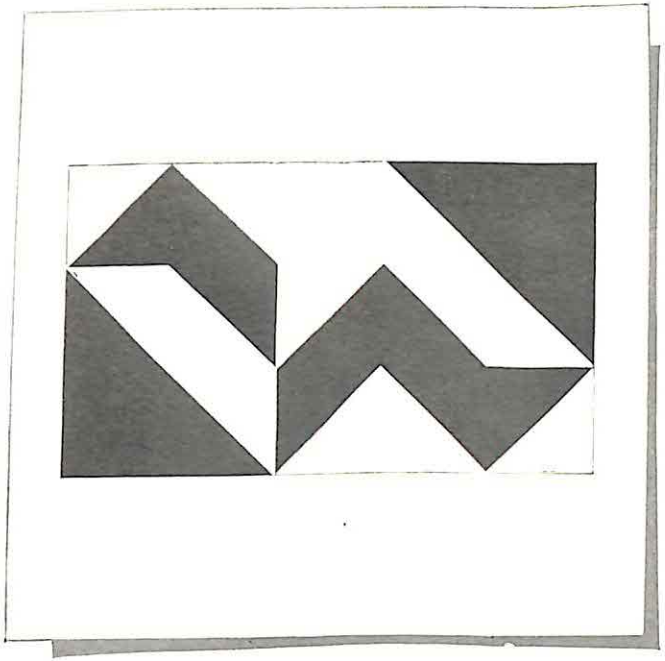


FIG. 12b. Sample Design Card for 15-Block Rectangle Placed within a Frame on a Rectangular Card. Changes are introduced here in the number of blocks required and in the use of margins of unequal width.

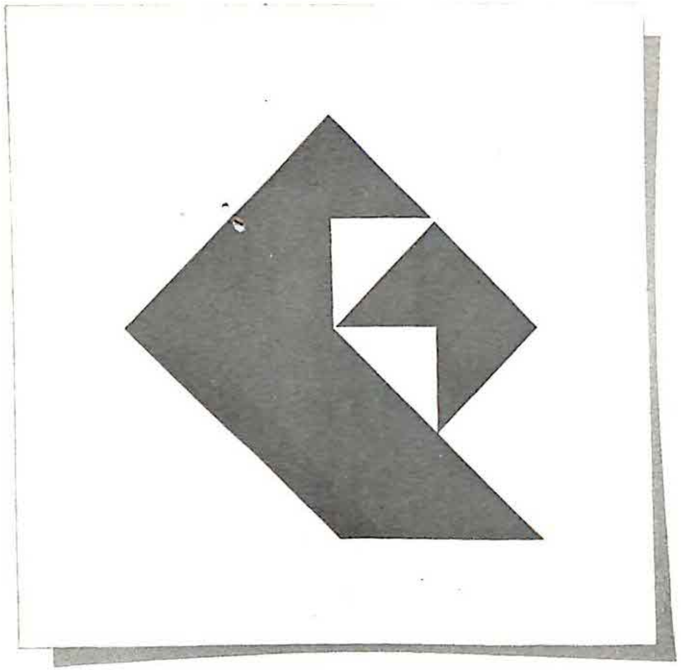


FIG. 13a. Sample Design Card for 16-Block Square on a Square Card. Difficulties are introduced here by the absence of block divisions and of a limiting frame, and by the use of margins which merge with the block background.

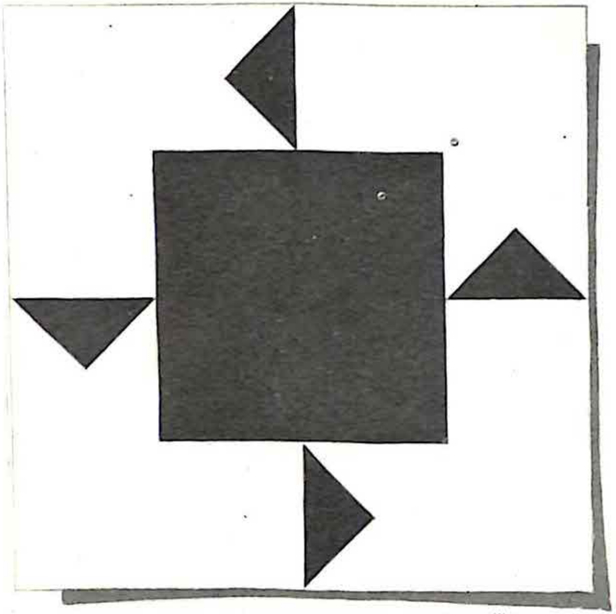


FIG. 13b. Sample Design Card for 16-Block Diamond Placed on a Square Card. Real difficulty is imposed here by the shape of the card, particularly since the design has no frame boundaries or block divisions.

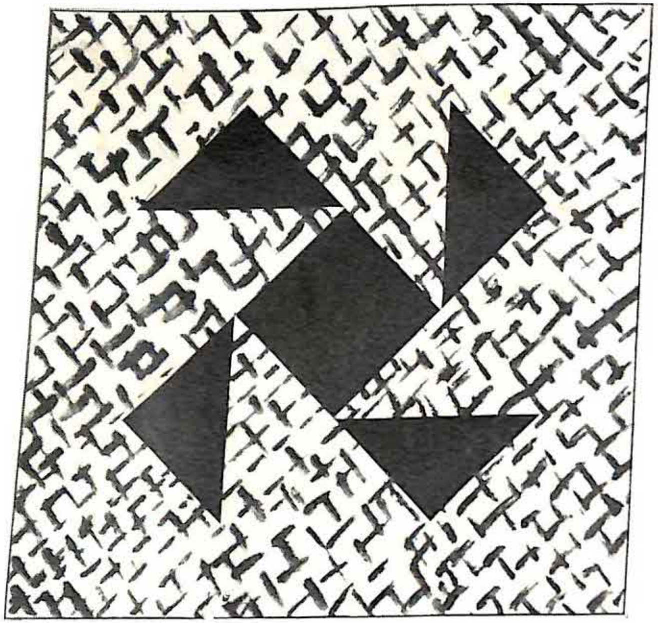


FIG. 14a. Sample Design Card for 16-Block Square on a Square Card with Structured Background. The absence of frame boundaries for the block construction and the use of margins and a closely structured background all impose additional demands on size perception and visual organization. In constructing this design there is a tendency to reproduce the central diamond by use of a single block.

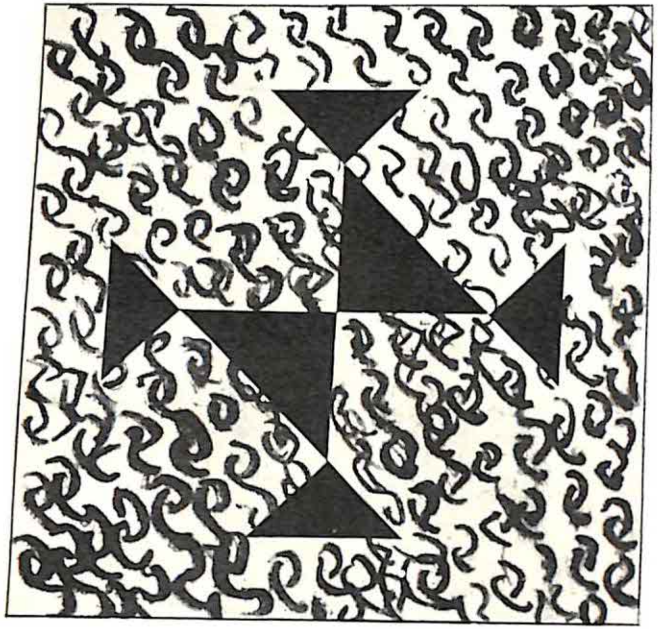


FIG. 14b. Sample Design, Card for 16-Block Diamond Placed on a Square Card with Structured Background. The competing background, the absence of a limiting frame, and the use of a diamond on a square all impose demands which measure skill and further clarify difficulties.

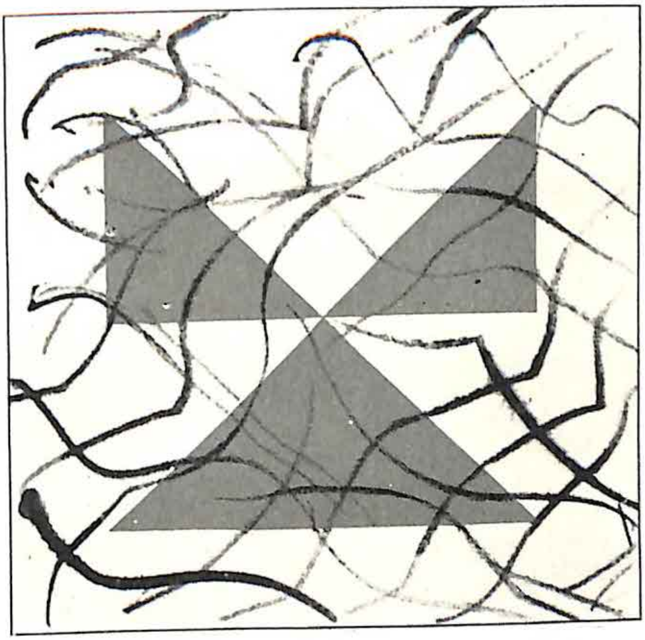


FIG. 14c. Sample Design Card for 16-Block Square Placed on a Square Card. This irregularly structured background introduces scattered points of distractibility and for some patients a factor of motion.

ances. This series of cards was first introduced to help in determining to what extent a patient was able to withstand the distraction of disturbing details within the background field. It was soon evident, however, that this kind of material was more broadly useful for both diagnostic and training purposes. For example, it was found that, in addition to the gross loss of visual orientation and organization resulting from figure-background competition, backgrounds of this sort frequently altered spatial orientation and perception (Fig. 14a and b). In some instances the fluid merging of design-background with extraneous margin seems to be a disrupting factor that destroys clues. In others, the configuration of the figure may be accurate, but distorted in size. The most common observable tendencies in such instances are the substitution of a single whole block for a four-block central square or diamond (Fig. 14a) and the substitution of a single block for a two-block triangle. In some cases it is noted that designs which had caused no difficulty when presented on unstructured backgrounds at earlier stages of training seemed to cause real distortions in size perception and spatial orientation when presented in a structured background. It was noted, furthermore, that to some patients some of these backgrounds caused more difficulty because they seemed to introduce an element of movement (Fig. 14c). Response to these designs, therefore, has not only been useful in uncovering previously unrecognized field abnormalities, but also in providing clues to locating practical visual-perceptual difficulties that are aggravated by motion or apparent motion. In this connection it is interesting to note that some of these patients have not only improved in visual perception and organization within the training situations which were planned with an awareness of the individual difficulties involved, but have been able to progress from block construction to vocational and sheltered workshop training situations which presented the need to apply the basic training in visual organization to real movements and distractions within the central and peripheral fields of vision. It is encouraging also to note that some of these patients have moved forward from controlled training situations to complex life situations involving such adaptations to movement as the ability to walk alone on the sidewalk, to go shopping, and to use the subway independently.

Although all of these experimental designs are black and white, the color intensities vary somewhat in order to provide for such individual differences as that encountered in a young woman whose abnormal response to white was so sharp that the presence of white destroyed all visual organization. This was true whether the presenting situation was a design card with a white background, a friend wearing a white blouse, an expanse of white tablecloth, or a book with very white pages and broad margins. During the first several weeks of training with this patient it was necessary to shade backgrounds into a pale gray and to screen off from the designs any margins which were not an integral part of the block construction. For most patients, variations in color intensities and in dull and shiny finishes have been found useful as a further means of measuring actual gains in the ability to analyze and organize a problem of visual perception, as differentiated from a purely concrete improvement dependent upon maintaining uniformity in the appearance of the cards.

For a clear, sharp figure-background contrast the most satisfactory cards have been made on unruled white file cards. The figures have been filled in with a black indelible pencil moistened in water and used as a brush. This medium combines the advantages of a finer control than an amateur could achieve with a brush and a solidly fused figure which shows no stroke marks. A durable finish is then put on the cards by painting the entire surface with a coat of colorless nail polish or by spraying it with a plastic spray.

On the basis of growing experience it is becoming evident that some of the experimental designs appear to be of little use except as additional practice material. Others appear to be basic—that is, for patients without organic impairment in visual perception and organization these designs usually cause no difficulty, whereas in most patients who do have severe disturbances in these areas these designs serve as starting points for differential diagnosis. It seems possible that further experience with all modifications may result in several serial arrangements of designs selected in terms of their greatest usefulness. Each of these designs would be made up in all levels of difficulty, as previously described. It is not the writer's intention to suggest that these series should not be used interchangeably in terms of the individual needs of individual patients. To the contrary, it is felt that a flexible use of the design cards has value as a way of defining more clearly the nature of perceptual difficulties, as a measure of progress in retraining specific disturbances of visual perception, and as an index of a developing power in visual organization.

Modifications in the Working Environment

Another variable involved in these modifications is the environment in which training and retraining is carried on. Under ideal conditions the room to be used should be large enough to permit each patient to be seated far enough from others to reduce extraneous visual stimuli resulting from the proximity of personal belongings or the movements of other patients. The walls and floors, as well as the surface of the working table, should be of dull finish and of uniform brightness. A warm gray, without gloss, has been found to be well suited to bringing out the maximum figure-background contrasts in block designs. The essential furnishings should also be of dull finish, and of a brightness corresponding to that of the walls, so that the general effect will be a merged background without the sharp breaks in contour which are frequently very disturbing to the brain-injured person with visual disturbances. Either natural or artificial light should be used alone. At no time should a combination of artificial and natural light be used. The artificial lighting should provide for individual shaded lights that can be adjusted to meet individual needs. In addition to collapsible and adjustable inclined surfaces built into the central working table, there should be at least one small table which can be adjusted in both height and angle for use by patients whose physical disabilities preclude the use of an adjustable stool.

Although the writer has not had such a working environment,[1] her experiences suggest that much can be accomplished with minor alterations and a few pieces of equipment which can usually be purchased at relatively small cost. The use of a table of conference size, rather than smaller tables, is advisable, particularly in small rooms. By covering its entire surface with a medium gray linoleum, a neutral background which accentuates the black-white contrast of the block designs is provided. This has the added advantage of reducing the glare reflected by varnished or painted surfaces. An adjustable long-arm lamp, or its equivalent, which can be adjusted to the needs of individual patients should be substituted for a ceiling light. Other special equipment should include at least two adjustable inclined surfaces. These may be lap tables with the legs and lower ledges removed, the sizes to be determined by what is available. Relative to this it has been noted that many of the patients who find an inclined surface helpful prefer the higher of the two models currently in use (Fig. 3). It is thought that this may be because, when elevated, this model seems to screen out more of the interference from the next larger visual field—namely, the surrounding table, its distant edge, and what lies beyond. These adjustable inclined surfaces are covered with thin mats of gray sponge rubber to keep the blocks from slipping and to reduce further the visual difficulty of reconciling the edges of the incline with the visual field between it and the edges of the work table. The angle at which these inclines are adjusted is determined by the patient's feeling of what seems best to him. When a flat surface is more comfortable, the patient usually indicates this and materials are presented accordingly.

The use of an adjustable stool or chair increases the possibility of adapting working distances and heights to the needs of each patient. This increase of distance above the working surface has often seemed to provide the first experience that a patient has had in achieving visual fusion or a "feeling of depth" in designs. The reality of this experience is evident in such spontaneous expressions as the comment made by a patient who, as she stood casually watching the completion of a block-design at an adjoining table, suddenly exclaimed, "Oh, if I could only reach far enough to build that design from here. Now it really makes sense!" Similarly it has not been unusual for a patient to ask, "Do you mind if I stand while I work? I can see it so much better. It has depth when I look at it from here."

---

[1] At the present time such a room is under construction at the Institute for the Crippled and Disabled.

Diagnostic Use of Modified Block Designs

In qualitative psychological evaluation no block work is attempted until there has been some exploration of the patient's ability to recognize similarities and differences in simple geometric forms. Solid block forms, line drawings, and pictures are used, and note is made of any clues which seem to be useful to the patient in selecting like forms. It is noted further whether he is able to differentiate only when the model is in all ways like the ones from which he is to make his choice, or whether he is equally well able to select the correct shape regardless of whether the model is a solid form or a line drawing, or whether it is larger, smaller, or a different color from the model presented to him. Note is also made of whether the patient is able to give these forms differential names, or is able only to indicate nonverbal recognition of likenesses and differences. It should be stated here that in order to understand the nature of the perceptual disturbance, it is fundamentally necessary to try to understand the patient's use of such terms as "straight," "up," "down," "long," "short," "big," and "little." In this connection the writer recalls the unnecessary complications introduced during the early weeks of training a patient with severe distortions of direction because of the failure to clarify with the patient the term "straight line." In this instance the patient's use of the term was strictly correct. By a "straight line" she meant any line which was not "wavy," while to the writer a "straight line" was one which was horizontal or vertical, as opposed to oblique or diagonal. This need for clarification is apparent both in birth-injured patients and in those whose brain injury was sustained at some later time.

If preliminary work with a patient does not suggest gross inability to recognize similar forms, the patient is first tried on a symmetrical black and white 16-block design which extends to the edges of the card on which it is drawn. No lines to mark the block divisions are shown on this first card. If, however, there is reason to believe that there may be real difficulty in this area, or if there are indications of strong feelings of insecurity, the first design to be presented will be one on which the block divisions are shown. When there is reason to believe that there may be serious disturbances in this area, the first design will be presented in actual block representation rather than as a flat card drawing. The patient's method of reproducing the design is carefully noted. Subsequent evaluative work with block design is determined by his initial performance and by spontaneous comments which he may make as he works. An effort is made to simplify the demands of the task to a point where either successful construction is possible or where a basic difficulty is revealed. Similarly, the demands are increased to the point where visual organization breaks down.

It should be said that in the qualitative psychological evaluation of brain-injured individuals, the modified black and white blocks and original design cards are used to an extent determined by individual needs. This involves adaptability in simplifying tasks to meet needs for increased concreteness, as well as exploring an individual's ability to cope with problems which require increased function at levels approaching the abstract. It involves also an understanding of the facts that visual distortions of size, form, space, and direction, as well as such relationships as figure to background, and details to wholes, can interfere seriously with block-design performance of an individual who shows relatively little impairment in areas other than those involving visual perception.

A clear-cut illustration of this is the case of a young woman of 26. On the Wechsler-Bellevue Verbal Scale her intelligence was 109, with indications of much better potential. On the Wechsler-Bellevue Performance Scale her intelligence was 53. Qualitative psychological and ophthalmological evaluations revealed actual limitations in vision as well as severe disturbances in visual perception, visuomotor performance, and visual organization. In practical function these had interfered seriously with her vocational training in the simplest office practice routines. At practical levels, not dependent upon the use of visually perceived clues, there was no apparent impairment of abstract functions. In spite of the fact that she wrote slowly, she interpreted and wrote her interpretations of ten proverbs in just slightly more than 20 minutes. The abstract quality of her mental function in this area was high as may be judged from the following samples: "Don't wait to lock the barn until after the horse is gone," interpreted as, "Take precaution to avert disaster"; "You can't prevent the birds of sorrow from flying over your head, but you can keep them from building nests in your hair," interpreted as, "You can't prevent adversity, but you can prevent it from stifling you"; and "Fine words butter no parsnips," interpreted as, "Fine talk alone can't earn your living."

As previously noted, the nature of many perceptual distortions can be observed through the use of design cards. A discussion of the differential diagnostic possibilities of these observations is not within the scope of this report, but present experience suggests that such a discussion might be worth consideration at some later time.

On the basis of the qualitative psychological evaluation, any adjustments which seem to be indicated are made before systematic psychological retraining begins. When a question is one that requires further specialized medical evaluation, the patient is referred for an ophthalmological examination or for further neurological consideration in terms of specific questions that have arisen during the evaluation.

For example, one might cite the case of a young cerebral palsied woman of 19. On the Verbal Scale of the Wechsler-Bellevue her intelligence was 90, on the Performance Scale it was 43. Although it was the psychologist's impression that her verbal potential was "at least average and that her common sense was superior," she presented training problems in visual perception, eye-hand coordination, and visual organization. On the basis of the qualitative psychological evaluation it seemed unlikely that she could achieve binocular vision. The specific questions raised by this evaluation were presented to her ophthalmologist, and a prompt and cooperative report was received granting us full permission to concentrate upon the development of efficient monocular right-eyed vision, and to use other training techniques which seemed to offer promise. In this case the patient from the outset was spared a repetition of the failures that made her dread academic teaching. Within the first six months she had the satisfaction of being able to write legibly for the first time in her life.

Another illustration is the case of a young woman who, while serving in the army, suffered a left hemiplegia as the result of surgery to remove a brain tumor. She had been in the training class for office practice for some time before being referred for qualitative psychological evaluation. Recent progress had been slowed up by disturbing visual symptoms. Furthermore, although she appeared capable of using public transportation she persisted in saying that she could not do this unless someone was with her because she became very confused—particularly on her left side. It had been thought by her social worker that she might be unwilling to give up these symptoms for fear of losing some of her disability compensation. On the basis of an extended evaluation there seemed reason to believe that there was a real field limitation on the left, for both distance and near vision, and that all visual function showed more confusion in the presence of moving objects. A major question seemed to be whether training emphasis should be directed toward treating a hysterical field limitation or toward developing greater skill in turning her body to view objects on the left. These and other questions were presented to the ophthalmologist. He reported that field studies revealed a left homonymous hemianopsia with macula sparing. His reply confirmed the fact that this patient would have difficulty viewing objects in the field of vision on the left side for both distance and near unless she moved her head to the left, and that training emphasis in occupational therapy and psychological retraining should concentrate on increasing her speed in turning her head to view objects on the left. It was reported further that her visual field difficulties were greater when viewing moving objects than when viewing static objects, and that this visual confusion would cause difficulties in transportation which could to a great extent be overcome by a process of retraining. Such training was instituted and the woman was relieved from further pressure to undertake independent travel at that time.

Quite differently oriented were the training goals and techniques in the case of a severely disabled cerebral palsied man of 22 who also appeared to have limitations in the peripheral fields. The visual confusion was increased when judgments had to be made in oblique directions. Although this patient traveled by himself, his gait when out of doors, or in a room where there was traffic with which he was not familiar, was conspicuously more unsteady and groping than when in more visually circumscribed situations. On the basis of the questions raised by our own evaluation, the ophthalmologist reported that this man's problem was not a field limitation resulting from brain damage, but rather was the result of very acute vision for distant objects. Because of this such a multiplicity of stimuli was admitted that he was unable to cope with them adequately, particularly since his condition was complicated by nystagmoid jerks and severe spasticity.

With very few exceptions this practice of delaying ophthalmological referrals until after our qualitative psychological evaluation has been productive in securing much more differential reports and recommendations from ophthalmologists than were received when a patient was referred for such an examination prior to qualitative evaluation. As a result it has been possible in some cases to reduce a patient's difficulties at the outset of training by such adaptations as concentration on the development of effective monocular vision, or by securing refractive assistance in "fogging" distortions, or in understanding symptoms as they relate to the whole problem of visual "extinction" and inattention as recently re-emphasized by Bender (3). These phenomena must be looked for since the failure to recognize them may result in serious misinterpretations of bizarre and striking deviations from normal responses, both in tests of visual perception and in practical daily living.

Training Adaptations of Modified Materials

The application of the previously described modifications of block designs to training and retraining depends entirely upon as complete an understanding of all the factors contributing to the patient's difficulties in form-space perception and visual organization as is possible in the present state of knowledge. *Individualized training in terms of individualized goals* is the basis of the entire technique, and upon it rests any favorable results that may be reported.

During the early weeks of training, work on a strictly individual basis has been found mandatory. During later stages it has been found that working with several patients at a time can be very valuable. The time when a transition from individual to group work should be made can only be determined on an individual basis. Similarly, decisions as to when an individual is ready to use Design Cards, and in what order these should be presented in either diagnostic or training situations depend upon the teacher's understanding of the individuals and upon the ingenuity and

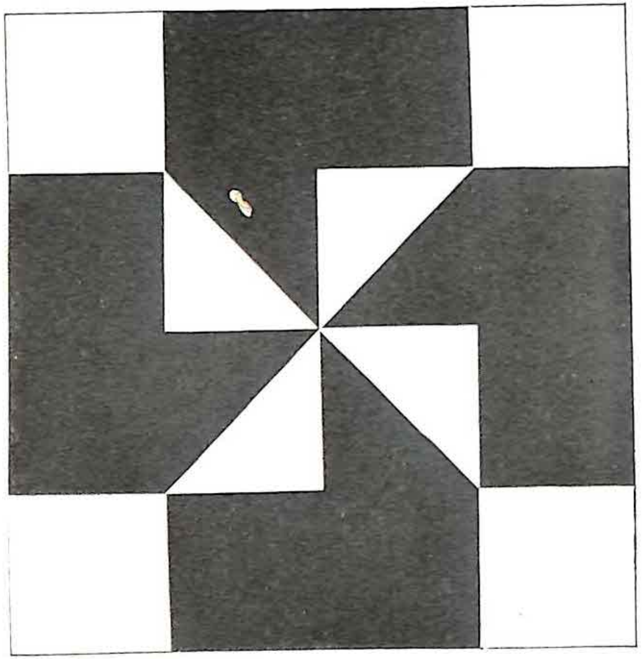


FIG. 15. Sample Design Card for 16-Block Square without Margins Which Combines the Use of 12 Solid-Color Blocks with 4 Diagonally Divided Ones.

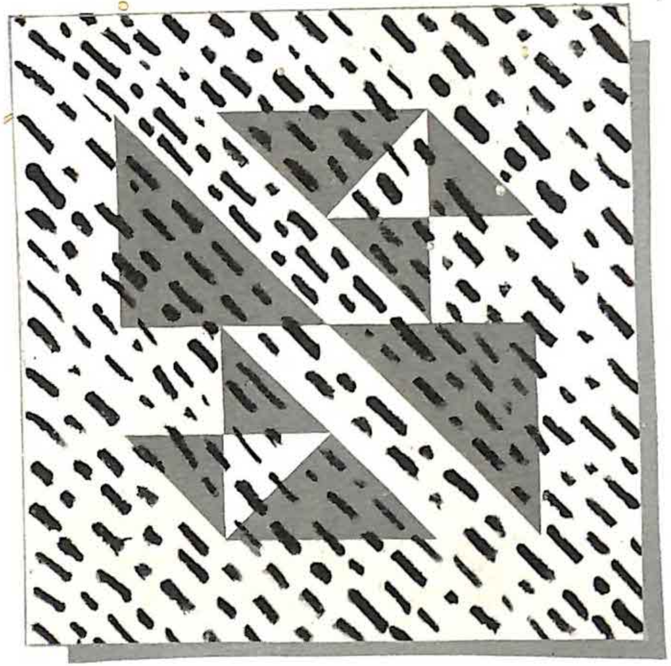


FIG. 16. Sample Design Card for 16-Block Square without Frame and with Margins Merging into Structured Background. Designs of this sort which combine several sizes of the same form have been useful at advanced levels of training.

skill which she can bring to the teaching situation. The fundamental guiding principle is that understanding must precede practice. To the brain-injured person, understanding comes through concrete experience at levels much below what an unimpaired person is apt to consider "rock-bottom" simplicity. Relative to this the writer has found that there seems to be little to lose and much to gain by beginning the training of all patients with various kinds of blocks and construction without the use of any design cards. For some patients this will mean a very brief experience with whole blocks in order to become familiar with their possibilities for constructing simple triangles and diamonds before they attempt to combine these in an organized way. From this exploratory experience the transition is made to building a block design from an actual block pattern. ,Only when a patient is readily able to note the likeness between his block construction and the block pattern is any attempt made to bridge the gap between a three-dimensional construction and its two-dimensional representation. For other patients the point of beginning will be solid block forms and training in getting a kinesthetic impression of the outlines of a circle, a square, a triangle, and a rectangle. For still others, the point of beginning may be a single block and enough half-blocks to build up a concrete understanding of each of the various sides of a whole block (Fig. 8 and 9). When necessary, the half-blocks can be placed directly on top of a whole block, thus matching each half-block with its mate. For patients with a minimum of manipulative difficulty, this use of half-blocks appears to be very helpful in providing some tangible clues to analyzing and organizing block designs at more advanced levels. For patients whose hand skills are too limited for this use of half-blocks, something of the same experience can be gained by watching the psychologist or another patient as he uses the half-blocks.

During the early stages with whole blocks and two-dimensional representations of them, the design cards are limited to those utilizing solid blocks or relatively simple geometric forms (Fig. 15). These triangles, diamonds, and three-quarter diamonds are used repeatedly, but in different positions, so that there is opportunity to develop real power in form perception and visual organization. Similarly, stripes are used in a variety of ways and directions. In later training, design cards are used that measure skill and provide training in size perception as well as in form. For example, designs planned for this purpose may be constructed of triangles of several sizes. Difficulty with size perception sometimes shows up very clearly when similar forms of different sizes are combined in a single design, particularly if the positions of the forms are varied (Fig. 16).

It should be noted here that the half-blocks were originally planned to make the visualization of a pattern apart from its immediate background more concrete. In actual use, these blocks have been helpful at various stages of improvement. In early training, before design cards are presented, the half-blocks are useful in providing concrete experience in the block construction of squares, triangles, diamonds, and diagonal stripes (Fig. 6 and 7). In this way each problem is more nearly reduced to uncluttered simplicity. At any level the half-blocks have been useful in providing a concrete understanding of size differences in forms of similar shape, as in triangles formed by one half-block, two half-blocks, four

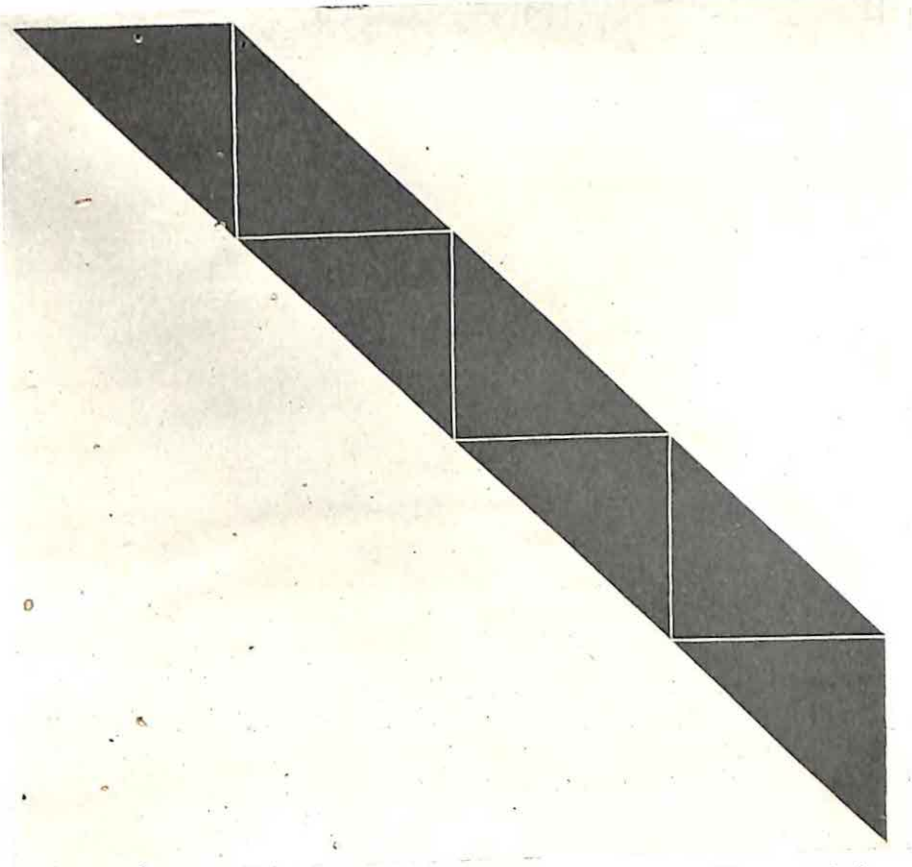


FIG. 17. The Use of 7 Black Half-Blocks to Construct a Diagonal Stripe Long Enough to Permit the Establishment of a Kinesthetic Impression of Its Direction.

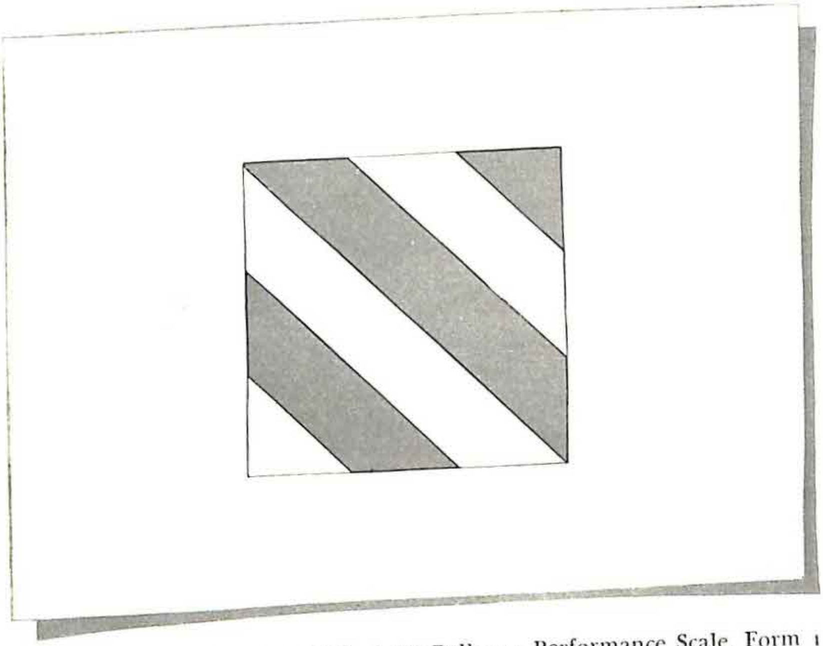


FIG. 18. Design Card 6 of Wechsler-Bellevue Performance Scale, Form 1 (Courtesy of Psychological Corporation, New York).

half-blocks, etc. More recently the half-blocks have been used to clarify the mechanical construction and direction of diagonal stripes. The feeling for a diagonal stripe, for example, appears to become more real when it is constructed by itself in a length sufficiently long to permit a patient to get a kinesthetic impression of it. This is possible when half-blocks are used and the block background does not have to be considered (Fig. 17). This appears to be especially important to brain-injured persons with spotty visual fields, who may be unable to appreciate a direction because too little of it can be seen in a design card such as Card 6 of the Wechsler-Bellevue Intelligence Scale, Form 1 (16) (Fig. 18). At all stages of training, the half-blocks have proved useful in providing a concrete experience as a basis for understanding the corners of squares and the tips of diamonds which are formed by a diagonally divided black and white block. In such instances the experience of being able to place a half-block in a position comparable to the card representation seems to help measurably in developing the ability to handle similar situations without the aid of blocks that are actually divided into tangible halves.

As analytical and organizational abilities develop, the abstract demands of tasks, as well as the physical complexity of the working environment, are increased proportionately. The patient no longer always works alone with the psychologist. He is brought into situations where he is obliged to cope with distractions from the corridor as well as with the movements and visual obstacles presented by having one or two other patients working at the same table or in close proximity. He is changed to various positions at the table so that practice is provided in adjusting his visual organization to the changes in his visual world. As skill and security increase, further demands are made upon a patient by increasing the difficulty of the construction itself. At times designs are constructed from memory. Sometimes they are reproduced in two-dimensional drawings, first from block models and later from memory. At other times they are constructed from memory, reversing the figure and background relationship by abstraction. Sometimes half-blocks are used to construct from memory design-figures independent of their block backgrounds. At some point in training the standard red-white, blue-yellow blocks and the Kohs Design Cards[5] are introduced. It has been interesting to note that at advanced levels of training, many patients do very well with this standard material, and to note the number of patients who without previous exposure to these materials since their initial experience with the Wechsler-Bellevue Scale are equally well able to construct the Kohs designs with the standard blocks, or to construct a Kohs design with black and white blocks, or to mentally convert one of the black and white experimental designs into a block construction in either the red-white or blue-yellow combination. In this connection it is perhaps noteworthy that the blue-yellow designs and the blue-yellow blocks seem to cause more difficulty with most patients than do the red and white. The relationship between this sort of block construction and the training possibilities in other aspects of abstract function would seem to be worth further exploration.

---

[5] The Kohs Design Cards, rather than the Wechsler-Bellevue or the Goldstein-Scheerer series, are used because this series is not used in the routine psychological evaluation in this situation.

Summary and Conclusions

On the basis of experience with the modified materials there appears to be sufficient evidence to support the following tentative conclusions:

1. Severe disturbances in the visual perception of form and space can be present in a brain-injured person simultaneously with good ability in the abstract thinking required to interpret fables and proverbs.

2. Impairments in abstract ability as tested by block designs are frequently complicated by visual-perceptual difficulties resulting from abnormalities in the visual fields, involuntary movements of the eyes or head, torsions, physical rigidities, and other conditions that provide the individual with visually distorted clues.

3. Great caution should be exercised in interpreting a brain-injured person's performance on standardized psychological tests, especially those involving picture arrangement, object assembly, and block design, and on projective techniques such as the Rorschach and Draw-a-Person tests. Responses to these tests may be markedly influenced by perceptual disturbances such as field abnormalities and distortions of size, space, distance, and of those lines and markings which serve as useful clues to persons without perceptual impairment. Although the distortions of color and movement and the confusions resulting from visual "extinction" are too little understood at the present time to be thoroughly evaluated, their importance should not be overlooked.

4. Performance with block designs may sometimes be appreciably improved when such factors as the angle of display, the position of the materials or of the individual, and distractions in the working environment are understood and measures are taken to reduce the resulting difficulties.

5. There is need to emphasize that disturbances in the perceptual areas of vision may be responsible for the underlying insecurity of the brain-injured individual and may be a valid basis of his hesitation and inadequacy in seeking or holding a job (4).

6. Training or retraining of visual perception and organization demands time sufficiently long to break down poor habits of visual perception and to develop new ones. Variations in training are necessarily dependent upon the needs and capabilities of individual patients. In some cases this will mean training that is limited to teaching an individual concrete ways to meet the production standards of a single task in a sheltered workshop. In other cases training may be undertaken at broader and more basic levels. In still others it may be found that the patient has already worked out fairly adequate ways of meeting the demands of his daily living and that any disturbance of his visual patterns would be unwise.

7. The emotional tensions of the brain-injured person whose perceptual disturbances do not respond to retraining can frequently be greatly relieved when a sincere attempt is made to help him and his family understand that the perceptual difficulties which block his performance or cause him to be misunderstood are respectable aspects of his disability. Following such a relief of tensions there has often been noted an increase in the patient's response to other forms of rehabilitation.

8. In our experience, the modification of the block design method described in this paper is one of the most useful single methods of evaluating disturbances of visual perception and organization. Further studies in this area would appear to be indicated.

REFERENCES

1. Bender, M. B., & Teuber, H. L. Spatial organization of visual perception following injury to the brain. *Arch. Neurol. Psychiat., Chicago,* 1947, **58,** 721-39.
2. Bender, M. B., & Teuber, H. L. Spatial organization of visual perception following injury to the brain. *Arch. Neurol. Psychiat., Chicago,* 1948, **59,** 39-64.
3. Bender, M. B. *Disorders in perception.* Publication No. 120, American Lecture Series. Springfield, Ill.: Charles C Thomas, 1952.
4. Frank, R. L. The influence of unrecognized difficulties. *Psychoanal. Quart.,* 1948, **17,** 84-96.
5. Gibson, J. J. *The perception of the visual world.* New York: Houghton Mifflin, 1950.
6. Goldstein, K., & Scheerer, M. Abstract and concrete behavior: an experimental study with special tests. *Psychol. Monogr.,* 1941, **53,** No. 2 (Whole No. 239).
7. Goldstein, K., & Scheerer, M. *Goldstein-Scheerer Cube Test.* New York: Psychological Corporation, 1945.
8. Goldstein, K. *Aftereffects of brain injuries in war.* New York: Grune & Stratton, 1942.
9. Hebb, D. O. *The organization of behavior: a neuropsychological theory.* New York: Wiley, 1949. Pp. 1-106.
10. Kestenbaum, A. *Clinical methods of neuro-ophthalmologic examination.* New York: Grune & Stratton, 1947. Pp. 23-52.
11. Kohs, S. H. *Design cards.* Chicago: Stoelting, 1923.
12. Lehtinen, Laura E., & Strauss, A. *Psychopathology and education of the brain-injured child.* New York: Grune & Stratton, 1948.
13. Sarason, S. B., & Potter, E. H. Color in the Rorschach and Kohs Block Designs. *J. consult. Psychol.,* 1947, **11,** 202-206.
14. Traquair, H. M. *An introduction to clinical perimetry.* (5th Ed.) St. Louis: Mosby, 1948. Pp. 38-39.
15. Werner, H., & Thuma, B. D. A deficiency in the perception of apparent motion in children with brain injury. *Amer. J. Psychol.,* 1942, **55,** 58-67.
16. Wechsler, D. *Block Design Cards. Wechsler Bellevue Intelligence Scale, Form 1.* New York: Psychological Corporation, 1948.

(Accepted for publication August 21, 1953)